# Characteristics of Genomic Mutations and Signaling Pathway Alterations in Thymic Epithelial Tumors

**Weiling Yang**
Sun Yat-sen University First Affiliated Hospital
**Sai Chen**
Sun Yat-sen University First Affiliated Hospital
**Xinxing Cheng**
Sun Yat-sen University First Affiliated Hospital
**Bo Xu**
Sun Yat-sen University First Affiliated Hospital
**Huilan Zeng**
Sun Yat-sen University First Affiliated Hospital
**Jianyong Zou**
Sun Yat-sen University First Affiliated Hospital
**Chunhua Su**
Sun Yat-sen University First Affiliated Hospital
**Zhenguang Chen** ( chzheng@mail.sysu.edu.cn )
Sun Yat-sen University First Affiliated Hospital https://orcid.org/0000-0001-6890-4063

**Research**

**Keywords:** thymic epithelial tumors (TETs), thymoma, PD-L1 expression, ErbB signaling pathway, MAPK signaling pathways

**Posted Date:** July 23rd, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-684468/v1



**Version of Record:** A version of this preprint was published at Annals of Translational Medicine on January 1st, 2021. See the published version at https://doi.org/10.21037/atm-21-5182.

# Abstract

**Purpose:** To elucidate mechanisms of thymic epithelial tumor (TET) canceration through characterization of genomic mutations and signal pathway alterations.

**Methods:** Primary tumor and blood samples were collected from 21 patients diagnosed with TETs (thymoma and thymic cancer), 15 of whom were screened by nucleic acid extraction and total exon sequencing. Bioinformatics was used to comprehensively analyze sequencing data for these samples, including differences in tumor mutation burden (TMB) and signaling pathways.

**Results:** We found that the gene with the highest mutation frequency in thymic carcinoma was *ZNF429* (36%). In addition, mutations in *BAP1* (14%), *ABI1* (7%), *BCL9L* (7%), *CHEK2* (7%) were only detected in thymic carcinoma, whereas *ZNF721* mutations (7%) were found only in thymoma. Mean TMB values for thymic carcinoma and thymoma groups were 0.722 and 0.663 mutations per megabase (Mb), respectively, differences that were not statistically significant. There were significant differences in enriched pathways for cellular components between tumor metastasis and non-metastatic samples. The ErbB signaling pathway was enriched in both the thymoma group and the intersection group, whereas “pathways in cancer” was found in both the thymoma group and thymic cancer group. In contrast, enrichment of longevity-regulating and MAPK signaling pathways was found only in the thymoma group.

**Conclusions:** We identified multiple differences in somatic genes and pathways, providing insights into differences between thymoma and thymic carcinoma that could aid in designing personalized clinical therapy.


# 1. Introduction

Thymic epithelial tumors (TETs) are the most common anterior mediastinal tumors in adults, accounting for about 0.2–1.5% of all malignant tumors.[1, 2] According to the histological classification of the World Health Organization (WHO), TETs can be divided into types A, AB, B1, B2, B3 and thymic carcinoma, of which thymic carcinoma is the most invasive subtype with the worst prognosis.[3] Generally speaking, thymoma is a tumor with local recurrence rather than metastasis. At present, the main treatment for thymoma is surgical resection with postoperative radiotherapy and chemotherapy.[4, 5] Thymic carcinoma, in contrast, is associated with a high risk of recurrence and death, despite surgery, chemotherapy and radiotherapy.[6] These differences are reflected in the corresponding 5-year median survival rates, which are 69% for thymoma but only 36% for thymic carcinoma.[7] Differences in molecular characteristics between thymoma and thymic carcinomas and changes in signaling pathways between thymoma and thymic carcinomas are currently unknown.

Currently, there are dozens of biomarkers related to checkpoint inhibitors, among which TMB (tumor mutational burden), PD-L1 (programmed cell death-1 [PD-1] ligand 1) and MSI/dMMR (microsatellite instability/deficient mismatch repair) have been verified by phase III clinical trials and are widely used in

clinical practice.[8–10] Although it has been reported that TMB-high (TMB-H) alone is not suitable for predicting immunotherapy effects against solid tumor types,[13] TMB is a biomarker for predicting PD-1/PD-L1 immune response.[11, 12] In adult tumors, the TMB of thymic tumors is lower, generally 0.48 mutations per megabase (Mb).[14] PD-L1 is commonly expressed in different TETs, including 23–92% of thymomas and 36–80% of thymic carcinomas.[15 – 21] A number of studies have investigated the relationship between PD-L1 expression and survival/prognosis of patients with TETs and have reported conflicting results. A study by Padda et al. showed that patients with high PD-L1 expression have poor overall survival.[15] However, Weissferdt et al. found no correlation between PD-L1 expression and prognosis.[21] Arbour et al. and Yokoyama et al. subsequently reported that patients with high expression of PD-L1 have a good prognosis.[18, 22] Studies have shown complex signaling pathway crossover in the tumor canceration process.[23] However, there are few studies on signaling pathways in the process of thymoma canceration.

To better understand the potential mechanisms and biological characteristics of TETs, we conducted next-generation sequencing (NGS) and signal pathway analyses on TET samples, analyzing somatic gene mutations, TMB, transition/transversion ratio (TI/TV), and mutated gene signaling pathways in thymoma and thymic carcinoma groups. Our findings could provide effective guidance for clinical intervention, thereby improving the long-term prognosis of patients with TETs.

# 2. Methods

## 2.1 Patient specimen acquisition.

A total of 21 patients with TETs were enrolled in this study, including 9 cases of thymoma and 12 cases of thymic carcinoma. PD-L1 was detected in all 21 patients with TETs, and full exon sequencing was performed in 15 patients. Slide-mounted tumor specimens were identified by hematoxylin and eosin (H&E)-staining according to the 2010 WHO classification of digestive system tumors. The included tumor tissue samples were histologically confirmed as adenocarcinoma by two molecular pathologists and shown to contain greater than 70% tumor cells. This study was conducted in accordance with the Helsinki Declaration, and the protocol was approved by the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University.[24] Prior to inclusion in the study, informed written consent and approval of the work was obtained from patients in all groups.

## 2.2 DNA extraction and qualification.

DNA was extracted from tissue sections of tumor samples using a DNeasy Blood and Tissue Kit (69504; Qiagen, Venlo, Netherlands) according to the manufacturer's instructions and characterized using an Agilent Bioanalyzer (Santa Clara, CA, USA). Germline mutations were determined by reference to matched leukocyte DNA.

## 2.3 Target gene sequencing and whole-exome sequencing (WES).

Libraries for the Illumina sequencing platform, which is equipped with a unique dual index (UDI) adapter that can mitigate label-skipping and label-mismatch problems, were prepared using a NadPrep DNA Library Preparation Kit (#1002101; Nanodigmbio, Nanjing, China). Hybridization capture-based target enrichment was performed using a NadPrep Hybrid Capture Reagents kit (#1005101; Nanodigmbio).

## 2.4 Variant annotation analysis and visual mapping.

False-positive results were excluded by realigning split-reads to hg19/GRch37 using Blast. The computational efficiency of WES data-capture was assessed using the Picard tool (http://broadinstitute.github.io/picard/command-line-overview. html#CollectHsMetrics), which filters low-quality bases, repetitive bases, bases that deviate from the target, and ends that overlap with reads at both ends owing to short inserts. This tool calculated the strictest depth distribution. For quality control purposes, sequencing data were defined as qualified if the mean bait coverage was greater than 100. Further local rearrangements were performed using SpeedSeq to improve the alignment of individual reads.[25] Somatic mutations were identified and indels were annotated using Mutect and Somatic Indel Detector software.[26, 27] Genetic variant data were annotated using ANNOVAR and Oncotator and converted to a MAF file using the maf tool.[28 – 30] Cancer-driver genes were analyzed using the Integrative OncoGenomics (IntOgen) platform, running Oncodrive FM and Oncodrive CLUST.[31] The landscape of the top driver mutation spectrum for tumors predicted by IntOgen, including mutation rate and mutation subclass/subtypes (Oncodrive FM *P*-value ≤ 0.1), was visualized using R Script.

## 2.5 Pathway enrichment analysis.

Wayne diagrams were used to illustrate differences in the frequency of somatic gene mutations between thymoma and thymic carcinoma groups because some gene mutations occurred only in one group, whereas others occurred in all groups. Gene Ontology (GO) canonical pathways associated with genes containing single nucleotide variations (SNVs) were analyzed using the R package.[32] *P*-values were calculated based on a hypergeometric distribution, and the false-discovery rate was corrected using the Benjamin-Hochberg method.[33]

## 2.6 Statistical analyses.

All correlations of clinical and biological variables were evaluated using the SPSS Statistics 22.0 package; where necessary, non-parametric tests (e.g., Welch's t-test and Wilcoxon signed-rank test) in the ggpubr module in the R package were used.[34, 35]

# 3. Results

## 3.1 Patient characteristics.

To better understand the molecular mechanisms underlying the genesis of TETs, we selected fresh frozen tumor tissues and corresponding blood samples from 15 patients with TETs from among the 21 enrolled study subjects and conducted bioinformatics analyses according to strict standards. Clinicopathological assessments indicated that 9 cases (S01, S04, S05, S06, S10, S12, S13, S14, and S15) were thymic carcinoma, of which one (S06) also had a lung metastasis. The remaining 12 cases (S02, S03, S07, S08, S09, S11, S16, S17, S18, S19, S20 and S21) were thymoma. Because the tumor cell content of samples S16, S17, S18, S19, S20, and S21 was less than 50%, the subsequent whole-exome sequencing step was not performed on these samples. Detailed clinical characteristics of patients are shown in Table 1. The PD-L1 positive rate was 33.3% (3/9) in thymic carcinoma and 58.3% (7/12) in thymoma (Table 1 and Fig. 1).

Table 1
Patient characteristics

| Characteristic | Number of cases | Proportion |
|---|---|---|
| Total number | *n* = 21 | |
| Age, years (mean) | 51 (21–69) | |
| Sex | | |
| Male | 15 | 71.4% |
| Female | 6 | 28.6% |
| Pathological type | | |
| Thymic carcinoma | 9 | 42.9% |
| Thymoma | 12 | 57.1% |
| PD-L1 positivity | | |
| Thymic Carcinoma | 3 | 33.3% |
| Thymoma | 7 | 58.3% |

# 3.2 Identification of somatic mutations in 15 patients with TETs.

We performed NGS-sequencing on DNA from 15 primary tumor tissues together with matched blood and annotated several somatic mutations using Mutect and Somatic Indel Detector. The average depth of whole-exome sequencing of tumor samples was 346.06X, and the average depth of sequencing of control samples was 127.96X. We detected a total of 1779 nonsynonymous single-nucleotide variations (SNVs): 912 in thymic carcinoma and 867 in thymoma (Supplementary Table 1).

As shown in Fig. 2, case S08 had the most SNVs, followed by case S03. Ranking the top 30 driver genes based on the frequency of their somatic mutation, we found that *ZNF429* (36%) was the gene with the highest mutation frequency. Missense mutations were the most common type of mutation, followed by nonsense, splice site, and other mutations. Among the top 30 driver genes, *BAP1* (14%), *ABI1* (7%), *BCL9L* (7%), and *CHEK2* (7%) were mutated only in thymic carcinoma, whereas *ZNF721* (7%) mutations were found only in thymoma.

We also calculated tumor mutation burden (TMB) using only somatic nonsynonymous mutations. On the whole, mean TMB values for thymic carcinoma and thymoma groups were 0.722 and 0.663 mutations per Mb, respectively, a difference that was not statistically significant (*P*-value = 0.817, Welch's t test; Fig. 3A). Notably, the TMB value for this study (0.698) was significantly higher than that reported in The Cancer Genome Atlas study of TETs (TCGA-THYM) (0.488) (*P*-value = 0.0038, Wilcoxon signed-rank test; Fig. 3B).

In principle, all types of mutations (substitutions, indels, rearrangements) and any accessory mutation characteristic, for example, the sequence context of the mutation or the transcriptional chain where the mutation occurred, can be incorporated into the set of features by which a mutational signature is defined. We extracted mutational signatures using base substitutions and assigned six classes of substitutions — C > A, C > G, C > T, T > C, T > A, and T > G — according to the pyrimidine of the mutated Watson–Crick base pair. We then compared the six mutation types with the TCGA-THYM database, and found that the proportions of these six mutation replacement types were roughly the same. The mutation percent of C > T was the highest among all substitutions and was significantly different compared with that for the TCGA-THYM database (*P*-value = 0.0014). T > G, C > A, and T > A substitutions were also significantly different between this study and the TCGA-THYM study (Fig. 3C, 3D).

## 3.3 Comparison of signaling pathway alterations between thymoma and thymic carcinoma.

To further characterize the functions of mutational genes and pathways involved in TET, we used the PANTHER classification system, an ontology-based pathway database coupled with data analysis tools.[36] Using pathological diagnosis results, we created a Venn diagram dividing driver mutant genes detected by whole-exome sequencing into three clusters: thymic carcinoma (38 genes), thymoma (38 genes), and the intersection of the two (15 genes) (Fig. 4A).

In thymic carcinoma, mutant somatic genes were mainly enriched for chronic myeloid leukemia (hsa05220), ErbB signaling pathway (hsa04012), and T-cell receptor signaling pathway (hsa04660) (Fig. 4B). By contrast, altered somatic genes in thymoma were mainly enriched for central carbon metabolism in cancer (hsa05230), acute myeloid leukemia (hsa05221), MAPK signaling pathways (hsa04010), and longevity-regulating pathways (hsa04213) (Fig. 4C). Shared driver gene enrichment pathways at the intersection of thymic carcinoma and thymoma clusters mainly included lysine degradation (hsa00310), endocrine and other factor-regulated calcium reabsorption (hsa04961), long-

term depression (hsa04730), long-term potentiation (hsa04720), gastric acid secretion (hsa04971), and the ErbB signaling pathway (hsa04012) (Fig. 4D).

# 4. Discussion

TET, a malignant tumor in the chest, is most common, albeit rare, in adult anterior mediastinal neoplasms, with an incidence rate of only 1.5–3.2 cases per 1000,000 people per year .[37] Among TETs, thymomas are often uniquely associated with autoimmune diseases, whereas thymic carcinoma exhibits more invasive characteristics in the clinic. A study by Radovich et al. reported a strong association of broad histological subtypes (A, AB, B, and thymic carcinoma) with multiple classes of aberrations that occur at different levels.[14] The etiology of TETs is unknown, largely owing to our limited knowledge of the genomic underpinnings of thymoma and thymic carcinoma.

Sequencing whole exons of 15 tumor-normal paired samples revealed some high-frequency mutation driver genes that were previously reported in studies of TETs, including *BAP1*, *ERBB4*, *HRAS* and *KIT*, among others.[38, 39] This analysis further showed that the highest frequencies of mutations were in *ZNF429* (36%) and *ZNF208* (29%) genes (Fig. 2). It has been reported that the *ZNF429* gene is significantly associated with acute postoperative pain in breast cancer, whereas *ZNF208* polymorphisms are associated with ischemic stroke in a southern Chinese Han population.[40, 41] Studies have found that PABPC1 (poly(A) binding protein cytoplasmic 1) is directly involved in the tumorigenesis of gastric cancer through promotion of cell proliferation; consistent with this, down-regulated expression of *PABPC1* is associated with tumor progression and a worse prognosis.[42, 43] Our analysis of the TCGA-THYM database further showed that the genes with the highest mutation frequency were *GTF2I* (50%), *HRAS* (8%), and *MUC16* (7%) (Fig. 5). It has been shown that mutated *GTF2I* is a thymoma-specific oncogene, and that cells expressing mutant GTF2I highly express genes involved in Wnt and Sonic hedgehog (SHH) signaling pathways.[14] Perplexingly, this latter study did not detect mutations in the *GTF2I* gene in thymoma or thymic carcinoma samples, possibly owing to insufficient sample size. MUC16, a mucin expressed at high levels on the surface of epithelial ovarian tumor cells, exerts effects similar to those of MUC1, which binds or aggregates neutrophils and other cell types to provide immune protection against tumors.[44, 45]

The expression of PD-L1 has been used as a tumor cell marker in a number of clinical trials and has been approved for this purpose.[46] It is thought that positive expression of PD-L1 in tumors is an indicator of patients who are more likely to respond to treatment.[47] However, some patients who test positive for PD-L1 may not respond to treatment; more importantly, some patients who test negative may still respond, making this an imperfect biomarker.[48] A number of studies have reported conflicting data on the relationship between PD-L1 expression and survival/poor prognosis in TETs.[15, 18, 21, 22] The rates of PD-L1 positivity in thymoma and thymic carcinoma in the current study (33.3% and 58.3%, respectively) are consistent with the reported positivity-rate range.[15 – 21]

High TMB is a characteristic associated with responsiveness to immunotherapy.[49] However, high TMB values (≥20 mutations per Mb) were found to be unsuitable for predicting immunotherapy effects in all types of solid tumors.[9] For different cancer types, establishing the high TMB threshold may require additional clinical studies and information statistics on a larger number of patients. We found that the TMB value in this study was significantly higher than that in the TGA-THYM study, suggesting that patients in our study would be more likely to benefit from immunotherapy (Fig. 3).

Because of differences between our findings regarding TMB and those reported by the TCGA-THYM study, we analyzed the types of mutation, defined by the TI/TV ratio. We found significant differences between our study and TCGA data with respect to C > A, C > T, T > A, and T > G mutations, but found no significant difference for C > G and T > C mutations (Fig. 3). Mutational signatures can be understood as different mutation processes that often generate different combinations of mutation types.[50] Thousands of somatic mutations can be identified in a single cancer sample, making it possible to decipher the mutant signature, even if several mutations are operative.[51] A previous study showed that the C > A mutational signature is associated with tobacco smoking, whereas C > T mutations are induced by ultraviolet light.[52]

Signal pathways are characterized by changes in multiple genes and expression levels, and usually involve simultaneous changes in multiple pathways, such as angiogenesis and notch signaling pathways. Cancers are not driven by single gene mutations or expression changes, but by coordinated changes affecting multiple signaling pathways.[53, 54] ErbB (erythroblastic oncogene B), also called the epidermal growth factor receptor (EGFR),[55] is generally expressed on the surface of epithelial cells, but is often overexpressed in certain tumor cells. Overexpression of EGFR is associated with tumor cell metastasis, invasion, and poor prognosis.[56] In the current study, we found enrichment of ErbB and T cell signaling pathways in a high proportion of samples in the thymic carcinoma group—pathways that were also found at the intersection of gene clusters (Fig. 4B, 4C). In the thymoma group, prominent pathways included longevity-regulating pathways (hsa04213) and central carbon metabolism in cancer (hsa05230) (Fig. 4C). Cell immortalization is an important stage in the tumorigenesis process.[57] The most reliable and repeatable longevity-promoting strategy for mammals is considered to be calorie restriction (CR). Among the pathways involved in regulating the CR effect are the insulin-like growth factor (IGF-1)/insulin signaling pathway, the sirtuin pathway, the adenosine monophosphate (AMP)-activated protein kinase (AMPK) pathway and the target of rapamycin (TOR) pathway.[58] It is thought that these pathways respond to CR via a common mechanism involving activation of autophagy, stress defense mechanisms and survival pathways, while weakening pro-inflammatory mediators and cell growth, thereby promoting cell health and, ultimately, longevity.[59] In the final analysis, the thymoma canceration process likely reflects benign or malignant mutations in multiple genes or changes in gene expression levels.

# 5. Conclusions

In present study, somatic gene mutations, tumor mutation burden, and TI/TV ratios in thymoma and thymic carcinoma employing bioinformatics were analyzed, and the differences between our findings and those of the TCGA project were also expounded. Dramatically, ZNF429, ZNF721, BAP1, ABI1, BCL9L, and CHEK2 can be found to varying degrees in thymoma or thymic carcinoma, respectively. Enrichment of longevity-regulating, MAPK, and ErbB signaling pathways still remain to be further improved. Multiple differences in somatic genes and pathways would aid in designing personalized clinical therapy in thymoma and thymic carcinoma.

# Declarations

### Data Availability Statement

All the related software and scripts are available from the corresponding author on reasonable request.


### Funding Statement

N/A

### Acknowledgements

We acknowledge the support of Shanghai Tongshu Biotechnology Co., Ltd.


### Authors' contributions

Conceived and designed the experiments: WY and ZC. Performed the experiments and acquired data: WY, SC, XC, BX, HZ, JZ, and CS. Analyzed the data: WY, XC, and ZC. Wrote the paper: WY, SC, and ZC. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

The use of human materials was approved by the Medical Ethical Committee of The First Affiliated Hospital, Sun Yat-sen University (Full name of the board/committee: the Medical Ethical Committee of The First Affiliated Hospital, Sun Yat-sen University).

### Consent for publication

Not applicable.

### Competing Interests

The authors declare that no competing interests exist.

# References

1. Jong, W.K.D., et al., Thymic epithelial tumours: a population-based study of the incidence, diagnostic procedures and therapy. European Journal of Cancer, 2008. **44**(1): p. 123-130.
2. Engels and A. Eric, Epidemiology of Thymoma and Associated Malignancies. Journal of Thoracic Oncology, 2010. **5**(10): p. S260-S265.
3. Best practices for the management of thymic epithelial tumors: A position paper by the Italian collaborative group for ThYmic MalignanciEs (TYME). Cancer Treatment Reviews, 2018.
4. Girard, N., et al., Thymoma: a focus on current therapeutic management. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer, 2009. **4**(1): p. 119-126.
5. Kelly, R.J., et al., Thymic malignancies: from clinical management to targeted therapies. Journal of Clinical Oncology, 2011. **29**(36): p. 4820-4827.
6. Eng, T.Y., et al., Thymic carcinoma: State of the art review. International Journal of Radiation Oncology Biology Physics, 2004. **59**(3): p. 654-664.
7. Scorsetti, M., et al., Thymoma and thymic carcinomas. Critical Reviews in Oncology/hematology, 2016. **99**(6): p. 332-350.
8. Yang, Q., et al., Multimodal detection of PD-L1: reasonable biomarkers for immune checkpoint inhibitor. American journal of cancer research, 2018. **8**(9): p. 1689.
9. Feng, L., et al., Whole-Exome Sequencing Characterized the Landscape of Somatic Mutations and Pathways in Colorectal Cancer Liver Metastasis. Journal of Oncology, 2019. **2019**.
10. Henriksen, A., et al., Checkpoint inhibitors in pancreatic cancer. Cancer treatment reviews, 2019.
11. Qin, B.-D., X.-D. Jiao, and Y.-S. Zang, Tumor mutation burden to tumor burden ratio and prediction of clinical benefit of anti-PD-1/PD-L1 immunotherapy. Medical hypotheses, 2018. **116**: p. 111-113.
12. Lu, S., et al., Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: a systematic review and meta-analysis. JAMA oncology, 2019. **5**(8): p. 1195-1204.
13. Xu, Z., et al., Assessment of tumor mutation burden calculation from gene panel sequencing data. OncoTargets and Therapy, 2019. **2019:12 3401–3409**.
14. Radovich, M., et al., The Integrated Genomic Landscape of Thymic Epithelial Tumors. Cancer cell, 2018. **33**(2): p. 244-258.e10.
15. Padda, et al., Diffuse High Intensity PD-1 Staining in Thymic Epithelial Tumors. Journal of Thoracic Oncology Official Publication of the International Association for the Study of Lung Cancer, 2015. **10**(3): p. 500-8.
16. Katsuya, Y., et al., Immunohistochemical status of PD-L1 in thymoma and thymic carcinoma. Lung Cancer, 2015. **88**(2): p. 154-159.
17. Yokoyama, S., et al., Clinicopathologic and Prognostic Implications of Programmed Death Ligand 1 Expression in Thymoma. Annals of Thoracic Surgery, 2016. **101**(4): p. 1361-1369.
18. Yokoyama, S., et al., Prognostic value of programmed death ligand 1 and programmed death 1 expression in thymic carcinoma. Clinical Cancer Research An Official Journal of the American

Association for Cancer Research, 2016. **22**(18).

19. Marchevsky, A.M. and A.E. Walts, PD-L1, PD-1, CD4, and CD8 expression in neoplastic and nonneoplastic thymus. Human pathology, 2016. **60**.
20. Higuchi, et al., PD-L1 Expression and Tumor-Infiltrating Lymphocytes in Thymic Epithelial Neoplasms. Journal of Clinical Medicine, 2019. **8**(11): p. 1833.
21. Wistuba, et al., Expression of PD-1 and PD-L1 in thymic epithelial neoplasms. Modern pathology, 2017.
22. Arbour, K.C., et al., Expression of PD-L1 and other immunotherapeutic targets in thymic epithelial tumors. Plos One, 2017. **12**(8).
23. Sanchez-Vega, F., et al., Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell, 2018. **173**(2): p. 321-337.e10.
24. Association, W.M., Declaration of Helsinki. Ethical principles for medical research involving human subjects. Journal of the Indian Medical Association, 2009. **14**(1): p. 233-238.
25. Chiang, C., et al., SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat Methods, 2015. **12**(10): p. 966-8.
26. Cibulskis, K., et al., Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol, 2013. **31**(3): p. 213-9.
27. DePristo, M.A., et al., A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet, 2011. **43**(5): p. 491-8.
28. Wang, K., M. Li, and H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res, 2010. **38**(16): p. e164.
29. Ramos, A.H., et al., Oncotator: cancer variant annotation tool. Hum Mutat, 2015. **36**(4): p. E2423-9.
30. Mayakonda, A., et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res, 2018. **28**(11): p. 1747-1756.
31. Gonzalez-Perez, A., et al., IntOGen-mutations identifies cancer drivers across tumor types. Nature methods, 2013. **10**(11): p. 1081-1082.
32. Yu, G., et al., clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 2012. **16**(5): p. 284-287.
33. Li, A. and R.F. Barber, Multiple testing with the structure-adaptive Benjamini–Hochberg algorithm. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2019. **81**(1): p. 45-74.
34. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. 2018; Available from: http://www.sthda.com/english/rpkgs/ggpubr.
35. Team, R.C. R: A Language and Environment for Statistical Computing. 2018; Available from: https://www.R-project.org/.
36. Huaiyu, M., et al., Large-scale gene function analysis with the PANTHER classification system. Nature Protocols, 2013. **8**(8): p. 1551-1566.

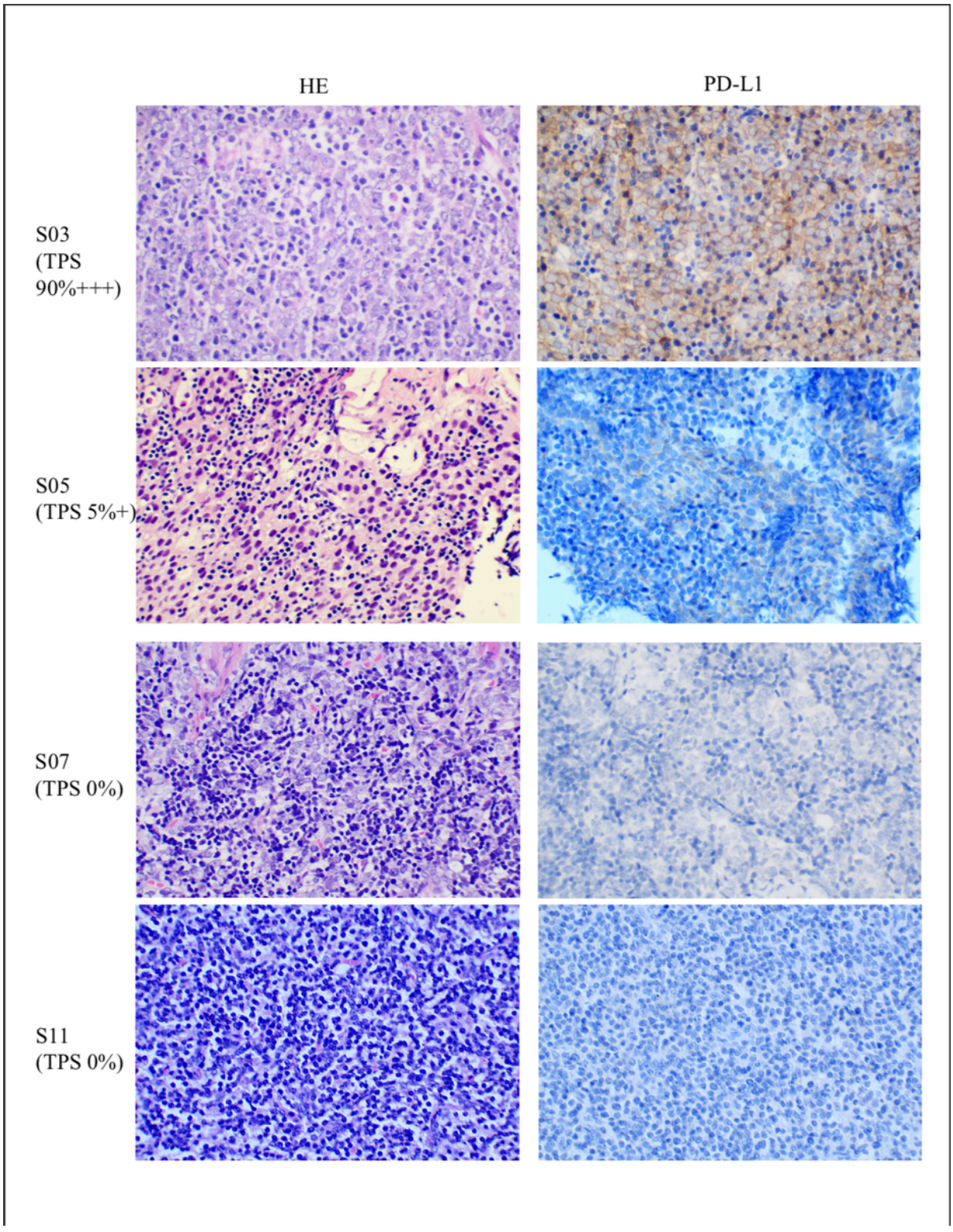
37. Janik, S., et al., Follistatin impacts Tumor Angiogenesis and Outcome in Thymic Epithelial Tumors. Scientific Reports, 2019. **9**(1).
38. Arenzana, T.L., et al., Tumor suppressor BAP1 is essential for thymic development and proliferative responses of T lymphocytes. Science Immunology, 2018. **3**(22): p. eaal1953.
39. Enkner, F., et al., Molecular Profiling of Thymoma and Thymic Carcinoma: Genetic Differences and Potential Novel Therapeutic Targets. Pathology & Oncology Research, 2017. **23**(3): p. 551-564.
40. Yu, J., et al., ZNF208 polymorphisms associated with ischemic stroke in a southern Chinese Han population. Journal of Gene Medicine, 2017. **19**(1-2): p. -.
41. Lee, E., et al., Genome-wide enriched pathway analysis of acute post-radiotherapy pain in breast cancer patients: a prospective cohort study. Human genomics, 2019. **13**(1).
42. Dong, H., et al., Long non-coding RNA SNHG14 induces trastuzumab resistance of breast cancer via regulating PABPC1 expression through H3K27 acetylation. Journal of Cellular and Molecular Medicine, 2018. **22**.
43. Yue-Qin, et al., PABPC1L depletion inhibits proliferation and migration via blockage of AKT pathway in human colorectal cancer cells. Oncology letters, 2019.
44. Sasamoto, N., et al., Development and validation of circulating CA125 prediction models in postmenopausal women. Journal of ovarian research, 2019. **12**(1): p. 116.
45. Fraschilla, I. and S. Pillai, Viewing Siglecs through the lens of tumor immunology. Immunological reviews, 2017. **276**(1): p. 178-191.
46. Chiarucci, C., et al., Circulating Levels of PD-L1 in Mesothelioma Patients from the NIBIT-MESO-1 Study: Correlation with Survival. Cancers, 2020. **12**(2).
47. Guo, M., et al., Expression and clinical significance of LAG-3, FGL1, PD-L1 and CD8T cells in hepatocellular carcinoma using multiplex quantitative analysis. Journal of translational medicine, 2020. **18**(1): p. 306.
48. Fang, W., et al., The Clinicopathological Features and Genetic Alterations in Epstein-Barr Virus-Associated Gastric Cancer Patients after Curative Surgery. Cancers, 2020. **12**(6).
49. Newell, F., et al., Whole-genome sequencing of acral melanoma reveals genomic complexity and diversity. Nature Communications, 2020. **11**(1).
50. Alexandrov, L.B., et al., Signatures of mutational processes in human cancer. Nature, 2013.
51. Teugels, E. and S.D. Brakeleer, An alternative model for (breast) cancer predisposition. Npj Breast Cancer, 2017. **3**(1): p. 13.
52. Alexandrov, L.B., et al., Mutational signatures associated with tobacco smoking in human cancer. Science, 2016. **354**(6312): p. 618-622.
53. R, L., et al., Deubiquitinase USP10 regulates Notch signaling in the endothelium. Science (New York, N.Y.), 2019. **364**(6436): p. 188-193.
54. Karchin, R., et al. IDENTIFICATION OF ABERRANT PATHWAY AND NETWORK ACTIVITY FROM HIGH-THROUGHPUT DATA. in Proceedings of the Pacific Symposium. 2012.

55. Deqin, M., et al., Somatic deletions of the polyA tract in the 3' untranslated region of epidermal growth factor receptor are common in microsatellite instability-high endometrial and colorectal carcinomas. Archives of Pathology & Laboratory Medicine, 2012. **136**(5): p. 510-6.
56. Comparison of the characteristics of uncommon epidermal growth factor receptor (EGFR) mutations and EGFRtyrosine kinase inhibitor treatment in patients with nonsmall cell lung cancer from different ethnic groups. Experimental and Therapeutic Medicine, 2020.
57. Yang, J.H., et al., Long-term persistent infection of HPV 16 E6 up-regulate SP1 and hTERT by inhibiting LKB1 in lung cancer cells. Plos One, 2017. **12**(8): p. e0182775.
58. Wang, Y., et al., Transcriptome analysis revealed potential mechanisms of differences in physiological stress responses between caged male and female magpies. BMC Genomics, 2019. **20**.
59. Barzilai, N., et al., The Critical Role of Metabolic Pathways in Aging. Diabetes, 2012. **61**(6): p. 1315-22.

# Supplementary

Supplementary Table 1 is not available with this version

# Figures

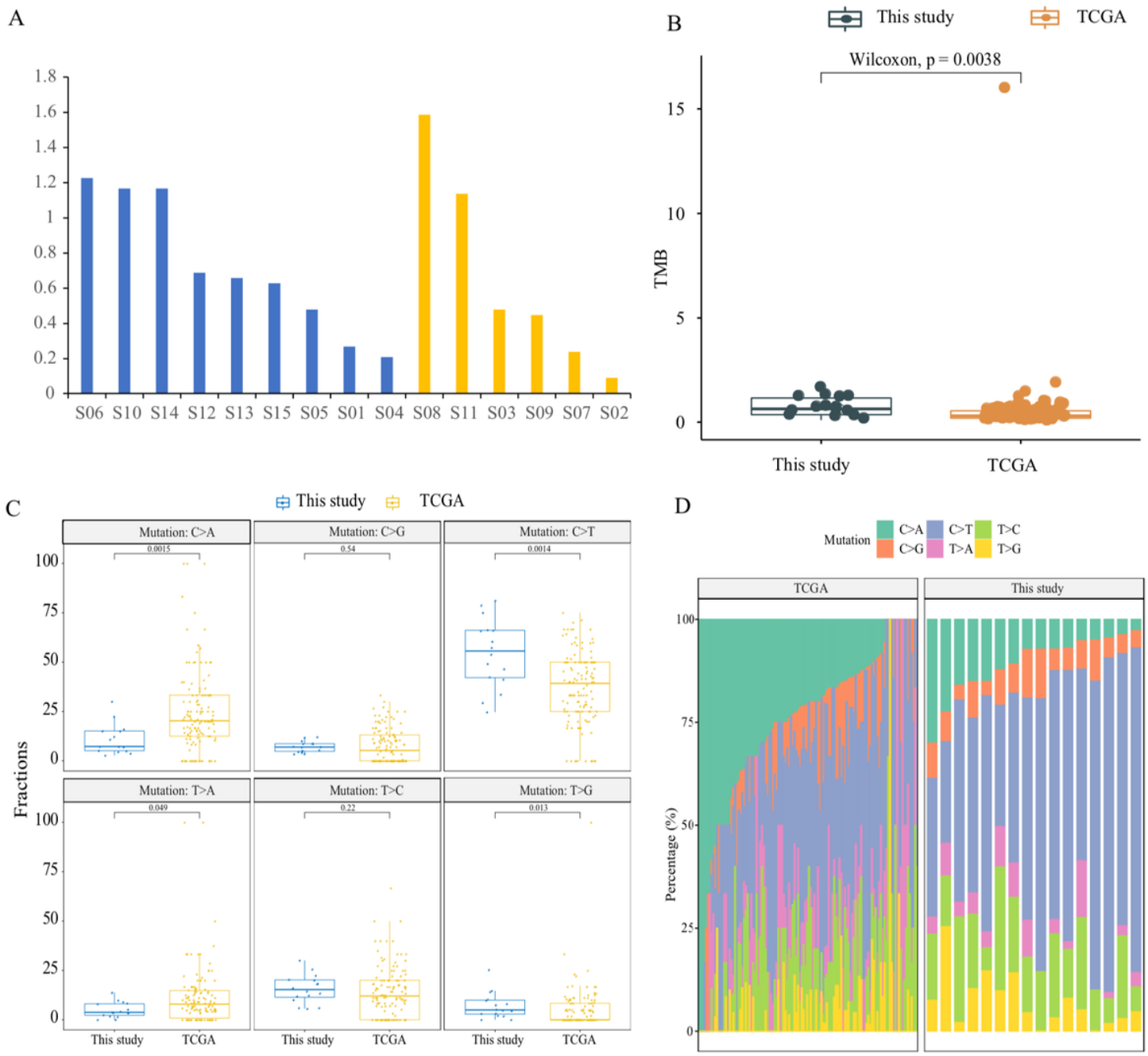



## Figure 1

Pictures of HE staining and PD-L1 detection result of thymic epithelial tumor samples (400x).

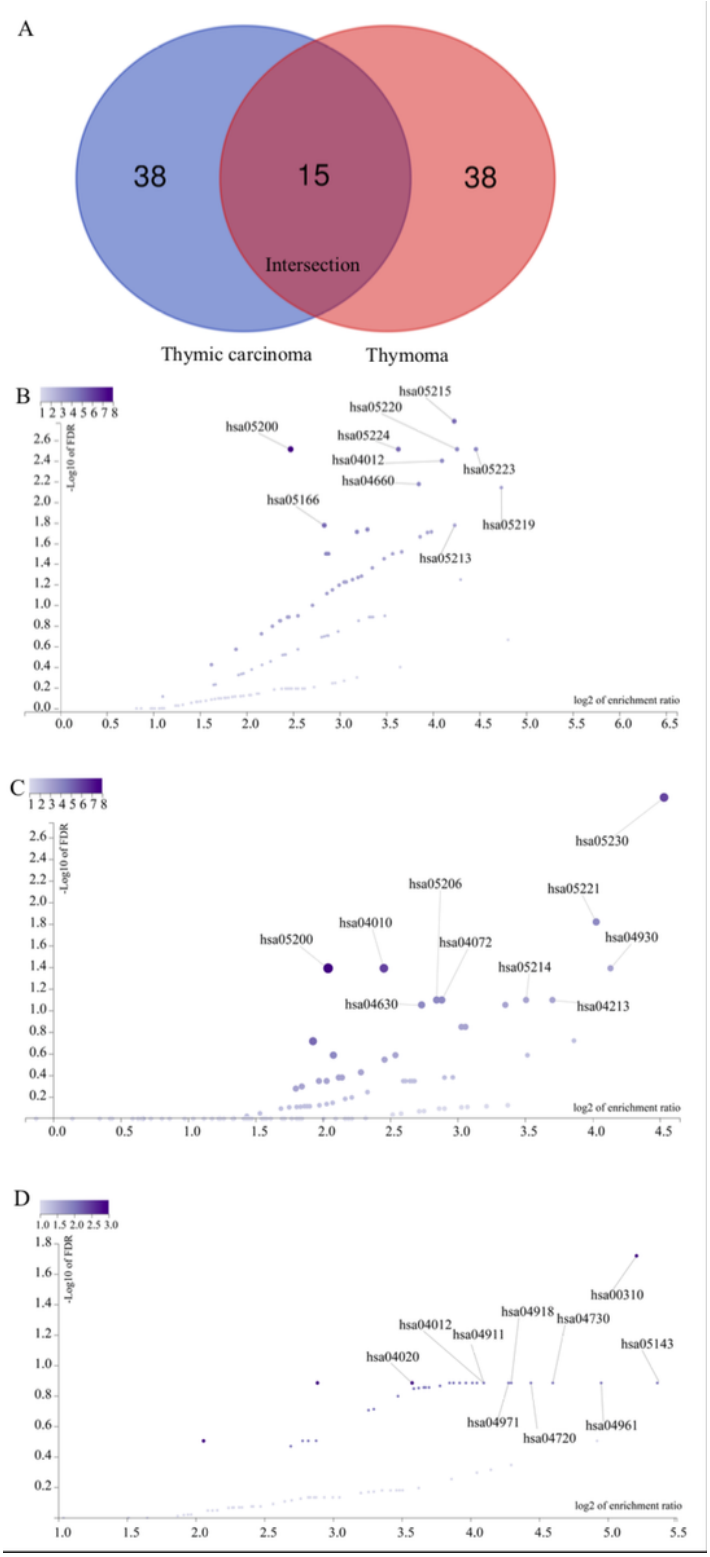

## Figure 2

Somatic mutations in thymic epithelial tumor (TET). Top: Variations in the numbers of mutations in each sample. Middle: Mutated genes and mutation types. Right: Frequency of mutations in somatic genes in TET samples.



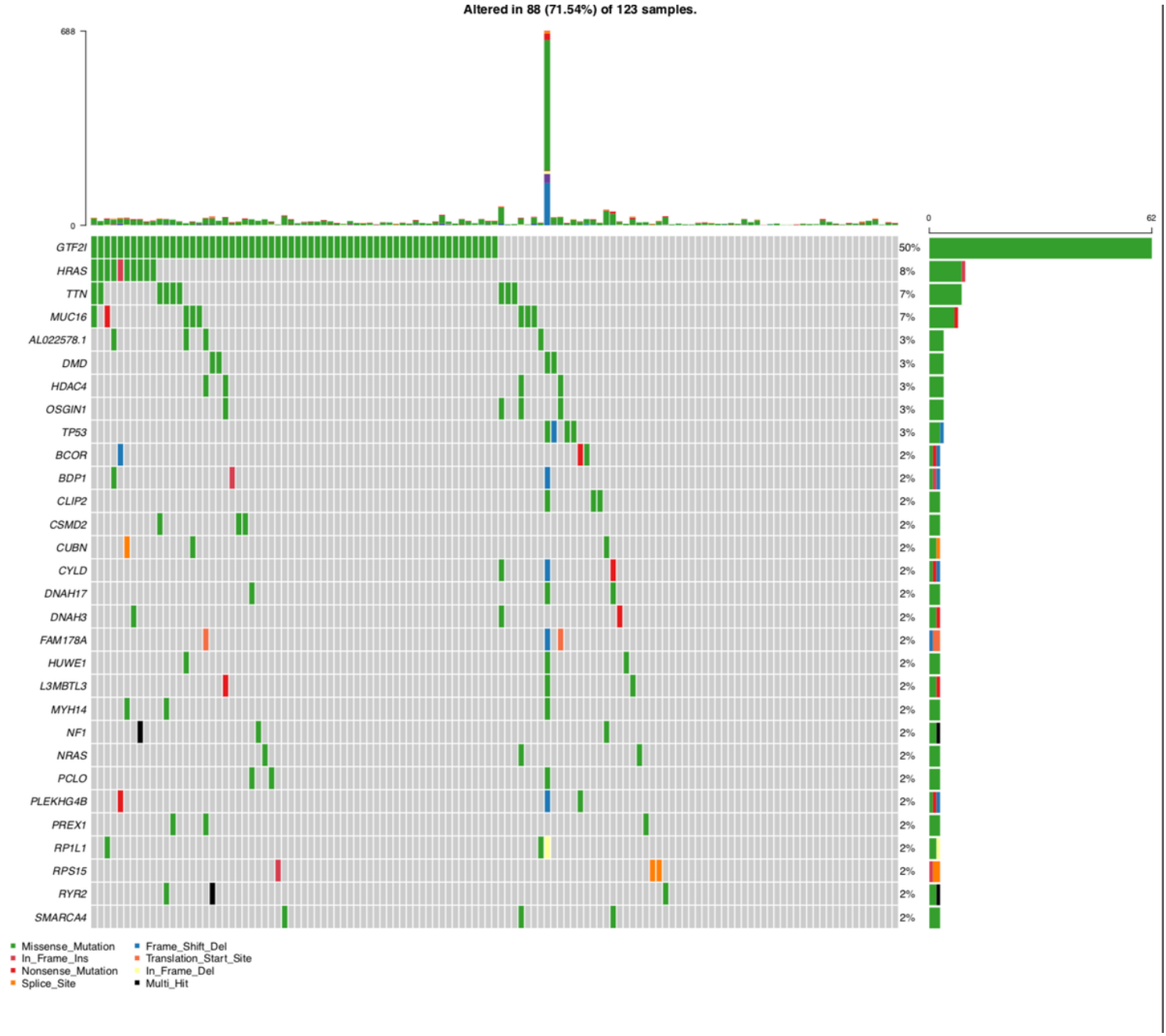

## Figure 3

Comparative analysis of tumor mutational burden (TMB) differences. A. TMB in thymic carcinoma and thymoma samples. B. Comparison of TMB between the data in this study and the TCGA-THYM study. C. Comparative analysis of the relative TI/TV ratio between this study and the TCGA-THYM study. D. Transition and transversion proportions shown for six nucleotide changes. The stacked proportion bar chart is sorted according to increasing TI/TV fraction.




## Figure 4

Analysis of pathways involving nonsynonymous genes in thymic epithelial tumor. A. Venn diagram showing gene distributions in three clusters. B. Volcano plot of uniquely mutated driver gene enrichment pathways in thymic carcinoma. C. Volcano plot of uniquely mutated driver gene enrichment pathways in thymoma. D. Volcano plot of shared driver gene enrichment pathways in the intersection group.




## Figure 5

Mutational landscape of somatic alterations in TCGA- THYM database.