

**Title: Genomic Profiling Identifies Somatic Mutations Predicting Thromboembolic Risk in Patients with Solid Tumors**

**Short Title: Genomics of Cancer-Associated Thrombosis**

**Word Count:** abstract; 238, text; 3,330

**Figure Count:** 3    **Table Count:** 2    **Reference Count:** 52

**Authors:** Andrew Dunbar, MD<sup>1</sup>, Kelly L. Bolton, MD, PhD<sup>1</sup>, Sean M. Devlin, PhD<sup>1</sup>, Francisco Sanchez-Vega, MD, PhD<sup>1</sup>, Jianjiong Gao, PhD<sup>1</sup>, Jodi V. Mones, MD<sup>1</sup>, Jonathan Wills, BS<sup>1</sup>, Daniel Kelly, MS<sup>1</sup>, Mirko Farina MS<sup>1</sup>, Keith Cordner BA<sup>1</sup>, Young Park MS<sup>1</sup>, Sirish Kishore, MD<sup>1</sup>, Krishna Juluru, MD<sup>1</sup>, Neil M. Iyengar, MD<sup>1</sup>, Ross L. Levine, MD<sup>1</sup>, Ahmet Zehir, MD, PhD<sup>1</sup>, Wungki Park, MD<sup>1</sup>, Alok A. Khorana, MD<sup>2</sup>, Gerald A. Soff, MD<sup>1</sup>, and Simon Mantha, MD, MPH<sup>1</sup>

**Affiliations:** <sup>1</sup>Memorial Sloan Kettering Cancer Center and <sup>2</sup>Cleveland Clinic

**Corresponding Author:** Simon Mantha, MD, MPH

Memorial Sloan Kettering Cancer Center

Koch Center, 545 East 73<sup>rd</sup> Street

New York, NY 10021

Phone: (646)608-3721

Fax: (929)321-8162

E-mail: manthas@mskcc.org

## Key Points

- Tumor mutations in *STK11*, *KRAS*, *CTNNB1*, *KEAP1*, *CDKN2B*, *MET* and *SETD2* modulate the risk of cancer-associated thrombosis.
- The presence of clonal hematopoiesis does not affect the risk of cancer-associated thrombosis.

## Abstract

Cancer-associated venous thromboembolism (CAT) is a well-described complication of cancer and a leading cause of death in cancer patients. The purpose of this study was to assess potential associations of molecular signatures with CAT, including tumor-specific mutations and the presence of clonal hematopoiesis. We analyzed deep-coverage targeted DNA-sequencing data of >14,000 solid tumor samples using the MSK-IMPACT™ platform to identify somatic alterations associated with VTE. Endpoint was defined as the first instance of cancer-associated pulmonary embolism and/or proximal/distal lower extremity deep vein thrombosis. Cause-specific Cox proportional hazards regression was used, adjusting for pertinent clinical covariates. Of 11,695 evaluable individuals, 72% had metastatic disease at time of IMPACT. Tumor-specific mutations in *KRAS* (HR=1.34 [1.09-1.64]; adjusted  $p=0.08$ ), *STK11* (HR=2.12 [1.55-2.89]; adjusted  $p<0.001$ ), *KEAP1* (HR=1.84 [1.21-2.79]; adjusted  $p=0.07$ ), *CTNNB1* (HR=1.73 [1.15-2.60]; adjusted  $p=0.09$ ), *CDKN2B* (HR= 1.45 [1.13-1.85], adjusted  $p=0.07$ ) and *MET* (HR=1.83 [1.15-2.92]; adjusted  $p=0.09$ ) were associated with a significantly increased risk of CAT independent of tumor type. Mutations in *SETD2* were associated with a decreased risk of CAT (HR=0.35 [0.16-0.79], adjusted  $p=0.09$ ). The presence of clonal hematopoiesis was not associated with an increased VTE rate. This is the first large-scale analysis to elucidate tumor-specific genomic events associated with CAT. Somatic tumor mutations of *STK11*, *KRAS*, *CTNNB1*, *KEAP1*, *CDKN2B* and *MET* were associated with an increased risk of VTE in solid tumor patients. Further analysis is needed to validate these findings and identify additional molecular signatures unique to individual tumor types.

## Introduction

Venous thromboembolism (VTE) is a frequent complication of cancer and a leading cause of cancer-related mortality.<sup>1,2</sup> Clinical risk factors known to portend an increased risk of cancer-associated venous thromboembolism (CAT) include body mass index (BMI), prior VTE events, inherited thrombophilias, blood cell counts, exposure to chemotherapy or recombinant erythropoietin, cancer stage, and the presence of underlying comorbid conditions, including infection.<sup>3-14</sup> These factors form the basis of risk assessment models frequently utilized to identify patients most likely to benefit from anticoagulant prophylaxis, most commonly the Khorana score.<sup>3,15</sup> However, despite improving CAT risk prediction, significant limitations remain with these models, and the optimal approach to pharmacological prophylaxis remains unclear. Identification of additional risk factors to enhance current CAT prediction models might help to better personalize VTE prevention and improve clinical outcomes for solid tumor patients.

The mechanisms driving the pathogenesis of CAT are likely complex and multifactorial and highlight the inextricably linked and highly dynamic interactions between the tumor, its microenvironment, and the hemostatic system. Known tumor-specific factors shown to promote venous thrombosis in cancer include: 1) over-expression and secretion of various pro-coagulant factors by the tumor, including tissue factor (TF) and TF-bearing microparticles; 2) activation of platelets and/or leukocytes by tumor secreted factors including pro-inflammatory cytokines; and/or 3) secondary effects of tumor cells on surrounding vasculature and tissue microenvironment as a result of aberrant pro-angiogenic growth factor stimulation.<sup>16</sup> Importantly, VTE rates vary significantly based on tumor sub-type, and even by histological sub-type within tumor types,<sup>17,18</sup> suggesting

that the individual mechanisms promoting thrombosis, and the degree to which the coagulation system is perturbed, are highly tumor-specific.

In this manner, cancer cell genotype is increasingly recognized as an important factor for VTE development.<sup>19</sup> Tumor-specific mutations in *KRAS*,<sup>20,21</sup> *ALK*<sup>22,23</sup> and *ROS1*<sup>24</sup> have been associated with an increased risk of thromboembolism. Results for *EGFR* mutational status and VTE risk have been mixed, with only one study showing an increased risk and others showing no effect or even a decreased risk of VTE.<sup>20,25-28</sup> Similarly, while many high-grade gliomas over-express TF in hypoxic conditions, the presence of an *IDH1* mutation appears to be protective against VTE in part due to reduced TF expression, further highlighting the influential role of tumor genotype in influencing CAT risk.<sup>29-32</sup> Also, evidence from an animal model suggests tumor-specific *MET* mutations might be associated with a hypercoagulable state.<sup>33</sup> Over-expression of TF and other pro-coagulant factors along with enhanced neoangiogenesis via constitutive VEGF stimulation are proposed mechanisms for the thrombophilic effect of oncogenic mutations.<sup>34</sup> Whether the pro-coagulant state of these tumors is simply a byproduct of aberrant signaling or serves directly to further promote tumor growth remains unclear; regardless, these data underscore the urgent need to identify the full extent of molecular events that might contribute to cancer-related thrombosis.

The emerging entity of clonal hematopoiesis (CH) is also increasingly recognized to influence thrombotic risk.<sup>35</sup> CH has been found to occur in >10-20% of otherwise healthy individuals over the age of 70 and in >25% of patients with solid tumor malignancies and is associated with adverse clinical outcomes.<sup>35,36</sup> Importantly, recent studies have implicated CH in an increased rate of arterial thrombotic events, including

myocardial infarction and stroke in otherwise healthy individuals.<sup>37</sup> In one retrospective series, JAK2 V617F mutant CH was also found to be associated with an increased risk of VTE and early onset atherothrombotic disease,<sup>38</sup> however, the role of CH influencing rates of VTE within the cancer patient population remains unknown.

While mutational profiling is increasingly used to inform important diagnostic, prognostic, and therapeutic considerations in solid tumor oncology patients, current clinically used assays are limited in the breadth and depth of detectable alterations. Since 2014, a custom hybridization capture-based next-generation sequencing assay known as IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets) has been utilized at Memorial Sloan Kettering Cancer Center (MSKCC) (MSK-IMPACT) to comprehensively profile patient tumors at the molecular level, with the capacity to detect somatic alterations in over 300 genes at a minimum depth of coverage of 91X.<sup>39</sup> To date, over 40,000 patient tumors have been profiled by MSK-IMPACT.<sup>40</sup> Given the availability of this platform and the unmet need of improving VTE prediction across the oncologic patient population, we sought to retrospectively analyze solid tumor MSK-IMPACT data to assess molecular determinants of CAT in patients treated at MSKCC.

## Methods

Patients and data capture: We included all adult patients who had solid tumor MSK-IMPACT testing between January 2014 and December 2016, as well as available tumor registry and electronic medical record (EMR) data (**Appendix A**). All patient samples and data were obtained following MSKCC institutional review board approval. Cancer-associated VTE was defined as any instance of pulmonary embolism (PE) or lower

extremity deep vein thrombosis (DVT), including both proximal and distal DVT events. Symptomatic and asymptomatic VTE episodes were included. Radiology reports from a list of studies susceptible to elicit a diagnosis of PE or lower extremity DVT (**Appendix B**) for the years 2011 to 2016 were searched for keywords indicative of VTE (**Appendix C**), and all positive reports were reviewed and verified by an observer (SM) for the presence of a VTE episode. The pre-IMPACT period from 2011-2013 was included to ensure that VTE episodes occurring in the years prior to cohort entry were identified. The R statistical platform was used along with package TM for the text mining portion of the work. Parallel to this search, pharmacy records were screened for mentions of therapeutic doses of an anticoagulant (**Appendix D**), and positive findings were reviewed by an observer (SM) for patients not already noted as a case in the radiology text search. All cases discovered in the radiology and anticoagulation search were reviewed and confirmed by a second observer (AD). Lastly, the corpus of medical notes ranging the years 2013 to 2016 was processed with a customized natural language processing (NLP) pipeline and data entry interface to identify missing lower extremity DVT and PE cases. Only events having occurred up to one year before cancer diagnosis were considered to be cancer-related; older episodes were recorded separately. Upper extremity DVT events were captured using the same customized NLP pipeline with each case reviewed by an observer (JM and SM). Information on tumor type and basic demographic characteristics was obtained from the institutional tumor registry and clinical database system. In order to facilitate the analysis and interpretation of the results, tumor type was simplified as detailed in **Appendix E**. Metastatic status during the study period was estimated by merging staging data from

the tumor registry and IMPACT records (**Appendix F**). Patients not marked as metastatic upon cohort entry had their clinical notes processed with the NLP pipeline to pick up any missed instances of metastatic disease. A manual audit was conducted on 300 randomly selected patients to assess the accuracy of data collection and aggregation.

MSK-IMPACT Sequencing: The MSK-IMPACT assay was performed as previously described.<sup>39</sup> Briefly, DNA from both tumors and patient-matched blood samples were obtained. Bar-coded libraries were then generated and sequenced. The custom targeted gene panel consisted of 341 genes (**Appendix G**). Subsequent versions of MSK-IMPACT contain additional numbers of genes (410 genes in Version 2; 468 in Version 3); however, these were ultimately excluded from the final analysis to maintain consistency across the cohort. Mean coverage across all tumor samples was 753X, with minimum depth of coverage of 91X. Raw results were run through a custom pipeline to identify somatic alterations. Germline mutations were filtered out. Final data from all samples were made publicly available online through the *cBioPortal for Cancer Genomics*.<sup>41,42</sup> For the current analysis, mutation data for all individuals were mapped as binary values (mutated vs unmutated). Only putative driver mutations were retained. Data on copy number alterations and gene fusions were also evaluated. In the case of fusions, only the gene exhibiting a potential change in expression was considered mutated. Fusions detected with the Archer panel (Illumina) were excluded, even though listed in *cBioPortal*, because only a small number of samples had this assay performed.



Statistical Analysis: The primary endpoint of time-to-CAT was defined as the time from accession to the minimum of CAT development, upper extremity DVT, death, or last follow-up in a one-year period following accession. CAT included lower extremity DVT (proximal or distal) and pulmonary embolism, whether symptomatic or not. Accession was defined as the date of blood sample receipt for IMPACT testing, and the approximate time a patient consents to testing. Cause-specific Cox proportional hazards regression estimated associations between select somatic mutations and the risk of developing a CAT episode. A separate model was built for each mutation; all models were adjusted for cytotoxic chemotherapy as a time-dependent covariate, history of VTE, current anticoagulation use, presence of metastatic disease, and age at accession. Additionally, all models were stratified based on tumor type, and because some tumor samples were previously banked, the years from the procedure to accession ([0], (0-0.25], (0.25-1], (1-5], [5,+)). Left truncation was used for the subset of procedures that occurred after accession but before the end of the one-year period. We only considered the 50 most commonly mutated genes in the cohort, in addition to *MET*, *ALK*, and *ROS1* which were selected based on prior data suggesting an effect on the risk of thromboembolism in cancer.<sup>22-24,33</sup> The Benjamini-Hochberg method was used to adjust *p*-values for false discovery. The predetermined statistical significance cutoff for the purpose of this analysis was 0.10. The R 3.6.1 statistical platform was used for multivariate analysis. For original data, please contact [manthas@mskcc.org](mailto:manthas@mskcc.org).

## Results

**Patient characteristics and incidence of VTE:** The schema used for VTE assessment in all patients is delineated in **Figure 1**. From 2014 to 2016, a total of 14,223 adult patients with solid tumor malignancy had MSK-IMPACT genomic sequencing and met initial inclusion criteria for this retrospective study. Of these, 1,962 were found to have at least one episode of CAT. 2,528 patients were excluded from cohort entry due to pre-existing CAT (n=1,104), upper extremity DVT (n=161) or because they had no clinical notes available in the EMR after the start of observation, consisting of the latest of IMPACT tissue sampling or blood sample accession. A total of 11,695 individuals were included in the final analysis, with 693 episodes of CAT recorded in the first year of observation. The highest event rates were in the sub-group of patients with pancreatic cancer (**Figure 2**). Characteristics of the patient cohort are presented in **Table 1**. The prevalence of individual cancer types included in the entire IMPACT cohort were generally reflective of the prevalence of cancer types within the larger oncologic patient population; however, breast cancer and prostate cancer, two of the most common cancer types, were under-represented, likely owing to fewer patients with these tumor types undergoing extended genomic testing during the observed time period. 163 patients (1.4%) had a VTE episode documented more than one year prior to cancer diagnosis. Importantly, most patients (72%; N=8,383) had metastatic disease at time of MSK-IMPACT.

Clinical risk factors were assessed for VTE risk. Consistent with previous reports, cytotoxic chemotherapy (HR 1.63; 95% CI 1.32-1.93,  $p<0.001$ ), prior VTE episode (HR 2.20; 95% CI 1.40-3.47,  $p=0.001$ ), and metastatic disease (HR 2.60; 95% CI 2.03-3.33,  $p<0.001$ ) were all strongly associated with CAT in multivariate analysis. Similarly, older

age, an elevated white blood count, a decreased hemoglobin, an increased platelet count and higher BMI values were associated with a higher risk of CAT. Data on those latter parameters were not available for all patients.

**Specific somatic alterations in tumor are associated with increased venous thromboembolism risk:** MSK-IMPACT data were then assessed to determine associations between individual genes and VTE risk across solid tumor malignancies. Tumor molecular profiles were analyzed across all cancer types and within sub-groups. Mutational frequencies across tumor types revealed rates of somatic alteration similar to those observed in previously published reports (**Supplementary Figure 1**).<sup>43</sup> *TP53* mutations had the highest prevalence, present in 42% of all tumor samples. *KRAS* and *EGFR* mutations were also frequent, found in up to 17% and 6% of tumor samples respectively and were identified across all cancer types, with a particularly high incidence in colorectal and pancreatic cancers. *IDH1* mutations were found to be enriched in glioma (24.9% of these patients), consistent with prior reports.<sup>44</sup>

There was no significant association between tumor mutational burden and risk of VTE (**Supplementary Figure 2**). There was no association of VTE risk with microsatellite instability (data not shown). Each gene was assessed in a separate regression model with the following covariates: age, previous VTE episode, anticoagulation, presence of metastatic disease and cytotoxic chemotherapy (time-dependent). Models were stratified based on tumor type and the time from a prior procedure to blood sample receipt for IMPACT germline control. Notably, mutations in *KRAS* (HR=1.34 [1.09-1.64]; adjusted  $p=0.08$ ), *STK11* (HR=2.12 [1.55-2.89]; adjusted

$p < 0.001$ ), *KEAP1* (HR=1.84 [1.21-2.79]; adjusted  $p=0.07$ ), *CTNNB1* (HR=1.73 [1.15-2.60]; adjusted  $p=0.09$ ), *CDKN2B* (HR= 1.45 [1.13-1.85]; adjusted  $p=0.07$ ), and *MET* (HR=1.83 [1.15-2.92]; adjusted  $p=0.09$ ) were found to be associated with a significantly increased risk of CAT independent of tumor type (**Table 2**). Mutations in *SETD2* were associated with a decreased risk of CAT (HR=0.35 [0.16-0.79]; adjusted  $p=0.09$ ). The unadjusted cumulative incidences of VTE for individuals with vs without a somatic mutation in these 7 genes are shown in **Figure 3**. Results stratified by tumor type are shown in **Supplementary Figure 3**. *CDK4* and *CDKN2A* mutations exhibited a trend towards an increased risk of CAT, with no statistically significant effect found after FDR-adjustment. A decreased risk of CAT was noted with *IDH1* mutations, but this effect was not significant after adjustment.

Given the observed effect of *STK11* mutations on the risk of CAT, a dedicated analysis was conducted on data from The Cancer Genome Atlas (TCGA) to evaluate the effect of *STK11* mutations on RNA expression for tissue factor (*F3*) and G-CSF (*CSF3*). Tissue factor is widely believed to be an important effector of CAT and has been shown to be upregulated by several oncogenes.<sup>34</sup> G-CSF has been demonstrated to be a likely mediator of the neutrophilia observed in lung cancer, also potentially being responsible for the formation of neutrophil extracellular traps and thromboembolism.<sup>45,46</sup> A dataset including 503 patients from TCGA and with lung adenocarcinoma was selected in *cBioPortal*, given the expected high prevalence of *STK11* mutations in this group.<sup>41,42,47</sup> RNA expression values were batch normalized (RNA Seq V2 RSEM) and a z-score threshold of +/- 2.0 was used. The mRNA expression z-scores relative to all samples were 0.58 vs -0.09 for *F3* (*STK11* mutated vs unmutated,  $p=1.334E-07$ ) and

0.39 vs -0.15 for *CSF3* (*STK11* mutated vs unmutated,  $p=9.175E-05$ ), as shown in **Supplementary Figure 4**. In order to assess the potential impact of increased transcription of G-CSF, a linear model predicting absolute neutrophil count based on the presence of a *STK11* mutation and adjusting for cancer type was fit using the MSK-IMPACT cohort. The presence of a *STK11* mutation was associated with an increase in 1,223 cells/mcL for the absolute neutrophil count ( $p<0.001$ ).

### **Clonal hematopoiesis in solid tumor patients does not increase the risk of VTE.**

We next assessed MSK-IMPACT data for the presence of CH and association with CAT. Of the 11,695 patients in the final cohort, 30% were found to have CH with a mutation in a known CH gene (**Supplementary Figure 5**). There was no significant association between any CH mutations, including *JAK2* V617F, and differential risk of CAT in the observed cohort (data not shown).

### **Discussion**

This work is the first broad search of a DNA tumor registry of its kind aimed to elucidate cancer-specific genomic determinants of CAT. Identifying mutations that might modify a patient's risk for CAT might not only improve prognostication, but also potentially uncover new pathophysiologic mechanisms of how aberrant signaling might lead to VTE development. The results of this study identified multiple individual tumor mutations influencing VTE risk in solid malignancy patients, including *KRAS* for which prior data exist.<sup>20,21,48</sup> Pre-clinical data demonstrate aberrant signaling through activating mutations of *MET* might confer a greater risk of CAT,<sup>33,49</sup> and this study suggests this

finding applies to human neoplasms. Additionally, *IDH1* mutations in gliomas have been demonstrated to be associated with a lower risk of VTE,<sup>31,32</sup> a finding supported by our analysis even though the effect did not reach statistical significance after adjustment for multiple comparisons. On the other hand, the increased risk of CAT demonstrated with mutated *STK11*, *CDKN2B*, *CTNNB1* and *KEAP1* genes has not been previously reported. However, the effect of *KEAP1* was mitigated in multi-gene regression models (data not shown), and mutations in this gene are strongly correlated with *STK11* defects<sup>50</sup> suggesting *KEAP1* mutations might not be an independent predictor of CAT risk. The mechanisms by which *STK11* mutations are associated with an increased risk of VTE are unclear, and might include increased neutrophil extracellular trap formation secondary to G-CSF production by the tumor. The possible explanations behind the decreased risk of CAT seen with *SETD2* mutations also remain to be determined.

Notably, previous studies revealed conflicting results in regard to *EGFR* mutational status in lung cancer and VTE risk, with either a protective effect, no effect or an increased risk.<sup>20,25-28</sup> In this large retrospective cohort including most solid tumor types, we failed to establish a significant association between CAT risk and the presence of an *EGFR* mutation. One possible explanation would be that in the last few years, patients with *EGFR*-mutated lung cancer have received *EGFR*-targeted therapies. In this regard, anti-*EGFR* monoclonal antibodies have been associated with an increased risk of arterial and venous thromboembolism, while *EGFR*-targeted tyrosine kinase inhibitors have not.<sup>51,52</sup> As such, differences in exposure to *EGFR*-directed therapies in different cohorts of patients with lung cancer could potentially result in discordant estimates of thromboembolism risk.

Available peripheral blood sequencing data for this large cancer cohort also allowed for the detection of CH and assessment of its effect on CAT risk. No significant association was found between the presence of any CH mutation and elevated VTE risk across all tumor types, including *JAK2* V617F, despite a previously published report of this association in non-cancer patient populations.<sup>38</sup> Germline mutations were also excluded from this analysis.

Any retrospective cohort study such as this is prone to several potential pitfalls in terms of data collection and analysis. Ascertainment of clinical events on a cohort of more than 10,000 individuals is particularly challenging. In the present case, only a fraction of the available clinical notes were reviewed by a human observer so it is possible that a small number of CAT events were missed. In addition, another important aspect in determining the validity of results for a clinical genomic study is the method of classifying genetic data with respect to mutational status. We decided to use a simple approach, labeling individual cohort members as being mutated or not for any given gene. One might devise a more sophisticated method and sub-classify participants based on the specific mutations encountered for the most commonly mutated genes like *KRAS*. Estimation of the false discovery rate (FDR) is another very important aspect of cancer genomic studies, and we chose 0.10 as a cutoff for significance. It is very possible that some genes with an FDR >0.10 in our models might be significant predictors of CAT. In this regard, any analysis measuring multiple associations requires special care to interpret results of statistical significance tests. How these mutations might contribute to increased thrombosis development is not known and will require further validation along with functional studies.

## **Conclusion**

Improved risk stratification methods for VTE risk are needed for the diagnosis and prevention of CAT. Enhanced genetic tools allow for the detection of molecular events that might contribute to CAT risk. Retrospective research such as the work presented here can help elucidate the pathophysiology of CAT and aid in the development of risk stratification tools, however the high dimensionality of the datasets used is challenging. The analysis we performed on a large cohort of patients with solid tumors suggests associations between cancer-specific mutations and CAT risk, justifying further validation of key results along with dedicated functional studies to better elucidate how tumor-specific alterations contribute to thrombotic disease in this high-risk population.

## **Acknowledgements**

The results published here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Memorial Sloan Kettering research support: NIH/NCI Cancer Center Support Grant P30 CA008748.

Dr. Dunbar receives support from the American Association of Cancer Research (17-40-11-DUNB)

Dr. Khorana acknowledges additional research support from the Sondra and Stephen Hardis Chair in Oncology Research and the National Heart, Lung and Blood Institute (U01HL143402).



Dr Park receives funding from the NIH (K12 CA184746 Paul Calabresi Award), the Parker Institute for Cancer Immunotherapy and SITC - Sparkathon TimIOs.

Dr Iyengar receives research support from the NCI/NIH (R01CA235711), the Breast Cancer Research Foundation and the American Cancer Society.

J.G. was supported by the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

### **Authorship Contributions**

S.M. planned the study and contributed to data collection, analysis and drafting of the manuscript.

A.D. drafted the manuscript and helped oversee the analysis.

S.M.D. performed the statistical analysis.

K.L.B. contributed to the analysis and drafting of the manuscript.

J.V.M., M.F., K.C. and Y.P. helped with data collection.

F.S.V. and J.G. contributed to the analysis.

K.J., J.W. and D.K. helped with data aggregation.

N.M.I. contributed to planning of the study.

R.L.L., S.K., A.Z., W.P., A.A.K. and G.A.S. contributed to drafting of the manuscript.

### **Disclosure of Conflicts of Interest**

S.K. serves as a consultant to Inari Medical. R.L.L. is on the supervisory board of Qiagen; is a scientific advisor to Loxo, Imago, C4 Therapeutics, and Isoplexis, each of which includes equity interest; receives research support from and consulted for

Celgene and Roche; received research support from Prelude Therapeutics; has consulted for Lilly, Janssen, Incyte, Novartis, and Gilead; and has received honoraria from Lilly and Amgen for invited lectures. W.P. receives funding from Merck, Astellas, and Gossamerbio, and is a consultant to Ipsen. A.K. received fees from Janssen, Bayer, Sanofi, Parexel, Halozyme, Pfizer, Seattle Genetics, Pharmacyclics, Pharmacyte, AngioDynamics, Leo Pharma, TriSalus and Medscape, along with grant support from Merck, Array, Bristol Myers Squibb and Leap. S.M. is principal owner of Daboia Consulting LLC. J.V.M. received fees from Sobi/Dova Pharmaceuticals. N.M.I. received consulting fees from Novartis and Seattle Genetics, along with grant support from Novartis.

## References

1. Khorana AA, Francis CW, Culakova E, Kuderer NM, Lyman GH. Thromboembolism is a leading cause of death in cancer patients receiving outpatient chemotherapy. *J Thromb Haemost.* 2007;5(3):632-634.
2. Horsted F, West J, Grainge MJ. Risk of venous thromboembolism in patients with cancer: a systematic review and meta-analysis. *PLoS Med.* 2012;9(7):e1001275.
3. Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. *Blood.* 2008;111(10):4902-4907.
4. Ay C, Dunkler D, Marosi C, et al. Prediction of venous thromboembolism in cancer patients. *Blood.* 2010;116(24):5377-5382.
5. Cronin-Fenton DP, Sondergaard F, Pedersen LA, et al. Hospitalisation for venous thromboembolism in cancer patients and the general population: a population-based cohort study in Denmark, 1997-2006. *Br J Cancer.* 2010;103(7):947-953.
6. Chew HK, Wun T, Harvey D, Zhou H, White RH. Incidence of venous thromboembolism and its effect on survival among patients with common cancers. *Arch Intern Med.* 2006;166(4):458-464.
7. Ahlbrecht J, Dickmann B, Ay C, et al. Tumor grade is associated with venous thromboembolism in patients with cancer: results from the Vienna Cancer and Thrombosis Study. *J Clin Oncol.* 2012;30(31):3870-3875.

8. Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ, 3rd. Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study. *Arch Intern Med.* 2000;160(6):809-815.
9. Starling N, Rao S, Cunningham D, et al. Thromboembolism in patients with advanced gastroesophageal cancer treated with anthracycline, platinum, and fluoropyrimidine combination chemotherapy: a report from the UK National Cancer Research Institute Upper Gastrointestinal Clinical Studies Group. *J Clin Oncol.* 2009;27(23):3786-3793.
10. Bohlius J, Wilson J, Seidenfeld J, et al. Recombinant human erythropoietins and cancer patients: updated meta-analysis of 57 studies including 9353 patients. *J Natl Cancer Inst.* 2006;98(10):708-714.
11. Bennett CL, Silver SM, Djulbegovic B, et al. Venous thromboembolism and mortality associated with recombinant erythropoietin and darbepoetin administration for the treatment of cancer-associated anemia. *JAMA.* 2008;299(8):914-924.
12. Nalluri SR, Chu D, Keresztes R, Zhu X, Wu S. Risk of venous thromboembolism with the angiogenesis inhibitor bevacizumab in cancer patients: a meta-analysis. *JAMA.* 2008;300(19):2277-2285.
13. Khorana AA, Francis CW, Culakova E, Kuderer NM, Lyman GH. Frequency, risk factors, and trends for venous thromboembolism among hospitalized cancer patients. *Cancer.* 2007;110(10):2339-2346.
14. Alcalay A, Wun T, Khatri V, et al. Venous thromboembolism in patients with colorectal cancer: incidence and effect on survival. *J Clin Oncol.* 2006;24(7):1112-1118.

15. Khorana AA, Francis CW. Risk prediction of cancer-associated thrombosis: Appraising the first decade and developing the future. *Thromb Res.* 2018;164 Suppl 1:S70-S76.
16. Falanga A SF, Russ L. Pathophysiology 1. Mechanisms of Thrombosis in Cancer Patients. In: G S, ed. *Thrombosis and Hemostasis in Cancer*. Vol. 179: Springer; 2019:11-36.
17. Chew HK, Davies AM, Wun T, Harvey D, Zhou H, White RH. The incidence of venous thromboembolism among patients with primary lung cancer. *J Thromb Haemost.* 2008;6(4):601-608.
18. Blom JW, Osanto S, Rosendaal FR. The risk of a venous thrombotic event in lung cancer patients: higher risk for adenocarcinoma than squamous cell carcinoma. *J Thromb Haemost.* 2004;2(10):1760-1765.
19. Magnus N, D'Asti E, Meehan B, Garnier D, Rak J. Oncogenes and the coagulation system--forces that modulate dormant and aggressive states in cancer. *Thromb Res.* 2014;133 Suppl 2:S1-9.
20. Corrales-Rodriguez L, Soulieres D, Weng X, Tehfe M, Florescu M, Blais N. Mutations in NSCLC and their link with lung cancer-associated thrombosis: a case-control study. *Thromb Res.* 2014;133(1):48-51.
21. Ades S, Kumar S, Alam M, et al. Tumor oncogene (KRAS) status and risk of venous thrombosis in patients with metastatic colorectal cancer. *J Thromb Haemost.* 2015;13(6):998-1003.

22. Verso M, Chiari R, Mosca S, et al. Incidence of Ct scan-detected pulmonary embolism in patients with oncogene-addicted, advanced lung adenocarcinoma. *Thromb Res.* 2015;136(5):924-927.
23. Zer A, Moskovitz M, Hwang DM, et al. ALK-Rearranged Non-Small-Cell Lung Cancer Is Associated With a High Rate of Venous Thromboembolism. *Clin Lung Cancer.* 2017;18(2):156-161.
24. Ng TL, Smith DE, Mushtaq R, et al. ROS1 Gene Rearrangements Are Associated With an Elevated Risk of Peridiagnosis Thromboembolic Events. *J Thorac Oncol.* 2019;14(4):596-605.
25. Wang J, Hu B, Li T, et al. The EGFR-rearranged adenocarcinoma is associated with a high rate of venous thromboembolism. *Ann Transl Med.* 2019;7(23):724.
26. Davidsson E, Murgia N, Ortiz-Villalón C, et al. Mutational status predicts the risk of thromboembolic events in lung adenocarcinoma. *Multidisciplinary Respiratory Medicine.* 2017;12(1).
27. Dou F, Li H, Zhu M, et al. Association between oncogenic status and risk of venous thromboembolism in patients with non-small cell lung cancer. *Respir Res.* 2018;19(1):88.
28. Berger N, Datta D, Tannenbaum S. Venous Thromboembolism and EGFR Mutation Status in Advanced Adenocarcinoma of Lung. CHEST Annual Meeting. Vol. 146; 2014.
29. Hamada K, Kuratsu J, Saitoh Y, Takeshima H, Nishi T, Ushio Y. Expression of tissue factor correlates with grade of malignancy in human glioma. *Cancer.* 1996;77(9):1877-1883.

30. Perry JR. Thromboembolic disease in patients with high-grade glioma. *Neuro Oncol.* 2012;14 Suppl 4:iv73-80.
31. Unruh D, Schwarze SR, Khoury L, et al. Mutant IDH1 and thrombosis in gliomas. *Acta Neuropathol.* 2016;132(6):917-930.
32. Mir Seyed Nazari P, Riedl J, Preusser M, et al. Combination of isocitrate dehydrogenase 1 (IDH1) mutation and podoplanin expression in brain tumors identifies patients at high or low risk of venous thromboembolism. *J Thromb Haemost.* 2018;16(6):1121-1127.
33. Boccaccio C, Sabatino G, Medico E, et al. The MET oncogene drives a genetic programme linking cancer to haemostasis. *Nature.* 2005;434(7031):396-400.
34. Tawil N, Bassawon R, Rak J. Oncogenes and Clotting Factors: The Emerging Role of Tumor Cell Genome and Epigenome in Cancer-Associated Thrombosis. *Semin Thromb Hemost.* 2019;45(4):373-384.
35. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. *Science.* 2019;366(6465).
36. Coombs CC, Zehir A, Devlin SM, et al. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell.* 2017;21(3):374-382 e374.
37. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-2498.
38. Wolach O, Sellar RS, Martinod K, et al. Increased neutrophil extracellular trap formation promotes thrombosis in myeloproliferative neoplasms. *Sci Transl Med.* 2018;10(436).

39. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J Mol Diagn*. 2015;17(3):251-264.
40. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703-713.
41. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.
42. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):p11.
43. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*. 2018;174(4):1034-1035.
44. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med*. 2009;360(8):765-773.
45. Tavakkoli M, Wilkins CR, Mones JV, Mauro MJ. A Novel Paradigm Between Leukocytosis, G-CSF Secretion, Neutrophil-to-Lymphocyte Ratio, Myeloid-Derived Suppressor Cells, and Prognosis in Non-small Cell Lung Cancer. *Front Oncol*. 2019;9:295.
46. Demers M, Krause DS, Schatzberg D, et al. Cancers predispose neutrophils to release extracellular DNA traps that contribute to cancer-associated thrombosis. *Proc Natl Acad Sci U S A*. 2012;109(32):13076-13081.



47. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018;173(2):291-304 e296.
48. Yu JL, May L, Lhotak V, et al. Oncogenic events regulate tissue factor expression in colorectal cancer cells: implications for tumor progression and angiogenesis. *Blood*. 2005;105(4):1734-1741.
49. Provencal M, Berger-Thibault N, Labbe D, et al. Tissue factor mediates the HGF/Met-induced anti-apoptotic pathway in DAOY medulloblastoma cells. *J Neurooncol*. 2010;97(3):365-372.
50. Arbour KC, Jordan E, Kim HR, et al. Effects of Co-occurring Genomic Alterations on Outcomes in Patients with KRAS-Mutant Non-Small Cell Lung Cancer. *Clin Cancer Res*. 2018;24(2):334-340.
51. Petrelli F, Cabiddu M, Borgonovo K, Barni S. Risk of venous and arterial thromboembolic events associated with anti-EGFR agents: a meta-analysis of randomized clinical trials. *Ann Oncol*. 2012;23(7):1672-1679.
52. Miroddi M, Sterrantino C, Simmonds M, et al. Systematic review and meta-analysis of the risk of severe and life-threatening thromboembolism in cancer patients receiving anti-EGFR monoclonal antibodies (cetuximab or panitumumab). *Int J Cancer*. 2016;139(10):2370-2380.

## Tables

		Overall N (%)	VTE Event N (%)
<b>Cytotoxic Chemotherapy*</b>	<b>No</b>	6759 (58%)	402 (58%)
	<b>Yes</b>	4936 (42%)	291 (42%)
<b>Age at Accession</b>	<b>Median (IQR)</b>	61 (52 - 70)	63 (54 - 71)
<b>Anticoagulant Subclass</b>	<b>Dabigatran</b>	35 (0%)	3 (0%)
	<b>Heparinoid</b>	135 (1%)	11 (2%)
	<b>None</b>	10925 (96%)	645 (95%)
	<b>Vitamin K Antagonist</b>	143 (1%)	10 (1%)
	<b>Xa-Direct Oral Anticoagulant</b>	178 (2%)	7 (1%)
<b>Metastatic Disease</b>	<b>No</b>	3312 (28%)	121 (17%)
	<b>Yes</b>	8383 (72%)	572 (83%)
<b>Tumor Type</b>	<b>Bladder</b>	400 (3%)	32 (5%)
	<b>Breast</b>	1690 (14%)	40 (6%)
	<b>Colorectal</b>	1084 (9%)	62 (9%)
	<b>Esophagogastric</b>	286 (2%)	30 (4%)
	<b>Gyn</b>	696 (6%)	51 (7%)
	<b>Head and neck</b>	327 (3%)	10 (1%)
	<b>Hepatobiliary</b>	383 (3%)	32 (5%)
	<b>High-grade glioma</b>	487 (4%)	44 (6%)
	<b>Lung</b>	1978 (17%)	161 (23%)
	<b>Melanoma</b>	470 (4%)	27 (4%)
	<b>Other</b>	1894 (16%)	87 (13%)
	<b>Pancreatic Adenocarcinoma</b>	487 (4%)	57 (8%)
	<b>Prostate Adenocarcinoma</b>	761 (7%)	24 (3%)
	<b>Renal</b>	325 (3%)	9 (1%)
	<b>Soft Tissue Sarcoma</b>	427 (4%)	27 (4%)
	<b>VTE Prior to Cancer Diagnosis</b>	<b>No</b>	11530 (99%)
<b>Yes</b>		165 (1%)	20 (3%)

**Table 1:** Characteristics of Patients

\* Cytotoxic chemotherapy anytime from day -30 to day 365 from accession date.

<b>Gene*</b>	<b>HR (95% CI)</b>	<b>p-value</b>	<b>FDR q-value</b>
<b>STK11</b>	2.12 (1.55 - 2.89)	0.0000023	<0.001
<b>CDKN2B</b>	1.45 (1.13 - 1.85)	0.0034719	0.072
<b>KEAP1</b>	1.84 (1.21 - 2.79)	0.0040963	0.072
<b>KRAS</b>	1.34 (1.09 - 1.64)	0.0058245	0.077
<b>SETD2</b>	0.35 (0.16 - 0.79)	0.0116198	0.088
<b>CTNNB1</b>	1.73 (1.15 - 2.60)	0.0089796	0.088
<b>MET</b>	1.83 (1.15 - 2.92)	0.0111671	0.088
<b>CDK4</b>	1.56 (1.06 - 2.29)	0.025	0.148
<b>IDH1</b>	0.45 (0.22 - 0.90)	0.024	0.148
<b>CDKN2A</b>	1.26 (1.02 - 1.55)	0.032	0.171
<b>SMAD4</b>	1.30 (0.96 - 1.76)	0.086	0.379
<b>MYC</b>	1.37 (0.96 - 1.94)	0.080	0.379
<b>TP53</b>	1.15 (0.98 - 1.35)	0.094	0.382
<b>FOXA1</b>	0.62 (0.33 - 1.16)	0.133	0.505
<b>APC</b>	0.74 (0.49 - 1.12)	0.159	0.561
<b>SMARCA4</b>	1.34 (0.87 - 2.05)	0.189	0.625

**Table 2:** Hazard Ratio of Venous Thromboembolism for Selected Cancer Somatic Mutations

\*Each gene evaluated in a separate Cox proportional hazards model adjusting for chemotherapy (time-dependent covariate), a history of previous venous thromboembolism, anticoagulation, presence of metastatic disease and age at accession. The model is stratified based on tumor type and the time from a prior procedure time to accession.

## Figure Legends

### **Figure 1:** Flow of Cancer-Associated Venous Thromboembolism Event Assessment

\*Patients who were followed actively at the medical center for the years 2014 to 2016.

CAT: cancer-associated venous thromboembolism.

NLP: natural language processing.

### **Figure 2:** One-Year Incidence of Cancer-Associated Venous Thromboembolism by Tumor Type

### **Figure 3:** One-Year Cumulative Incidence of Cancer-Associated Venous Thromboembolism by Mutation Type

### **Supplementary Figure 1:** Frequency of Mutations

### **Supplementary Figure 2:** Cumulative Incidence of Cancer-Associated Venous Thromboembolism by Tumor Mutational Burden for Most Frequently Observed Tumors

### **Supplementary Figure 3:** One-Year Cumulative Incidence of Cancer-Associated Venous Thromboembolism by Mutation Type Stratified by Tumor Type\*

\*Two tumor types with the highest mutation frequencies used for each gene

### **Supplementary Figure 4:** mRNA Expression Z-scores Relative to All Samples for F3 (A) and CSF3 (B) for Patients With vs Without a STK11 Mutation\*

\*Data from TCGA Pan-Cancer Atlas, all samples are from adenocarcinoma of the lung, n=503.

### **Supplementary Figure 5:** Cumulative Incidence of Cancer-Associated Venous Thromboembolism by Clonal Hematopoiesis Status

PanCan PD: pan-cancer, putative driver clonal hematopoiesis

CH: clonal hematopoiesis

Active\* patients aged  $\geq 18$  years with IMPACT blood control accession on or before 12-31-2016: n=14,223

Positive for automated medication search only: n=821

Positive for both automated medication and radiology record search: n=1,299

Positive for automated radiology record search only: n=667

Negative on manual review of radiology record: n=156

Confirmed by manual review of radiology record: n=1,143

Anticoagulation history reviewed manually: n=977

Additional patients found on review of VQ scans: n=4

Confirmed by manual review of radiology record: n=222

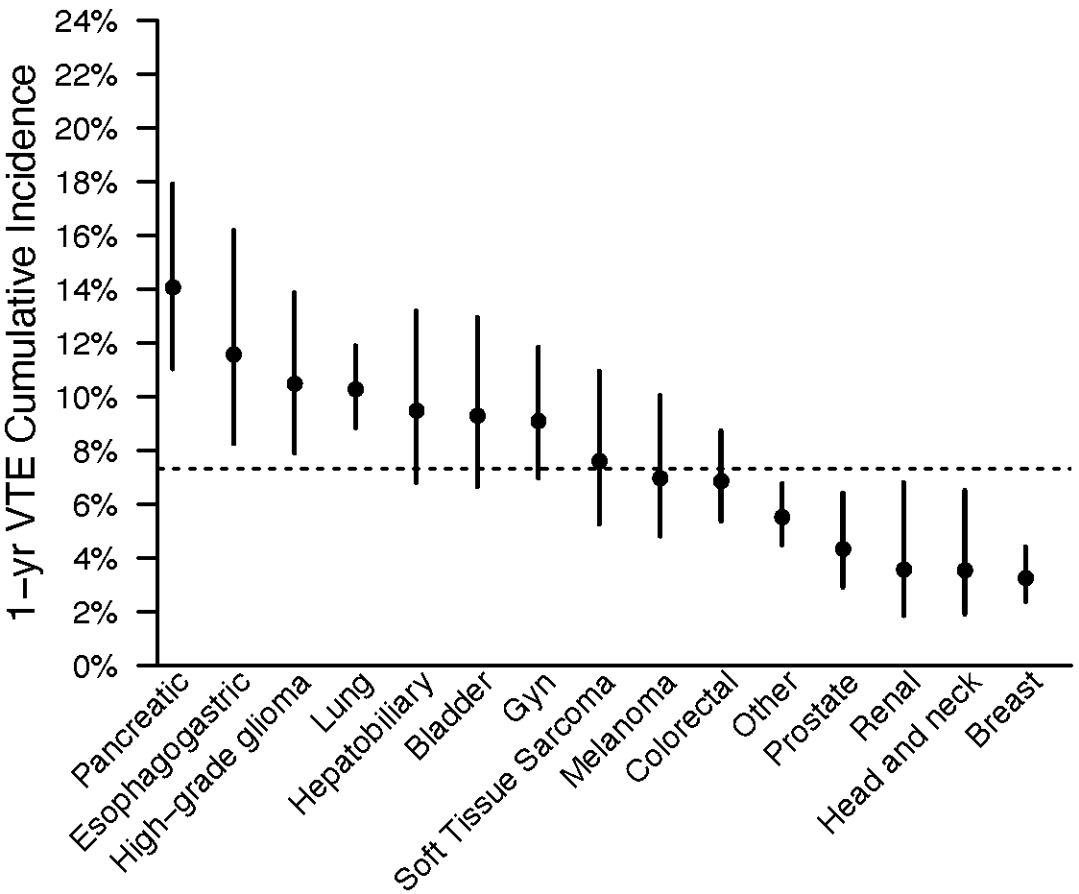
Confirmed by manual review of anticoagulation history: n=270

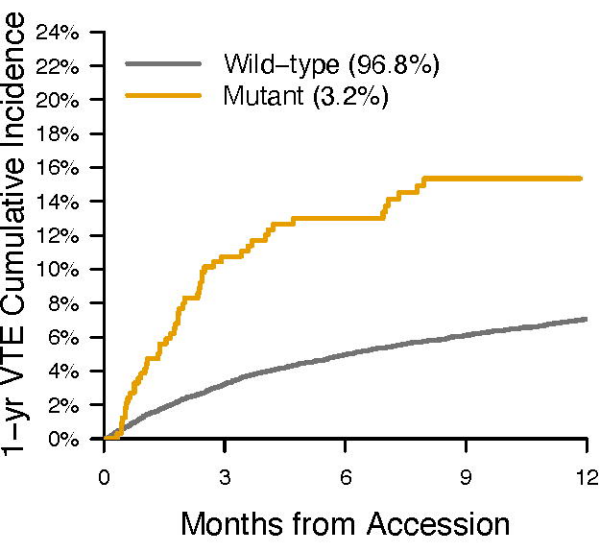
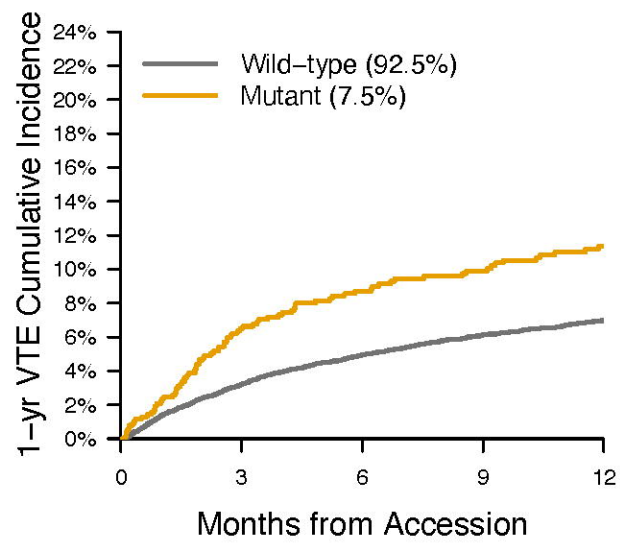
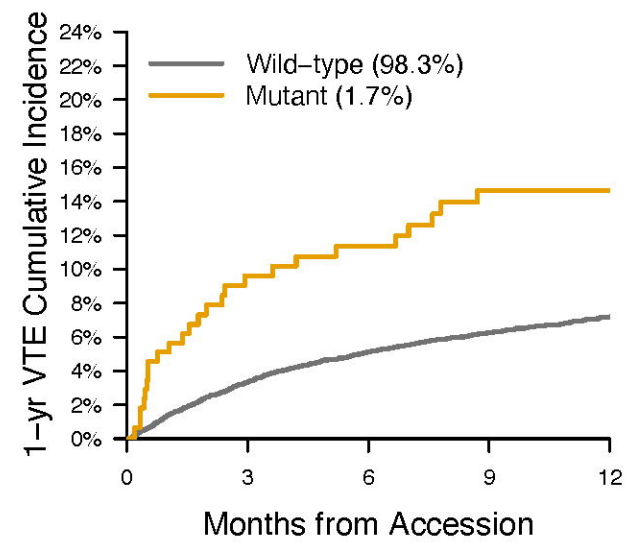
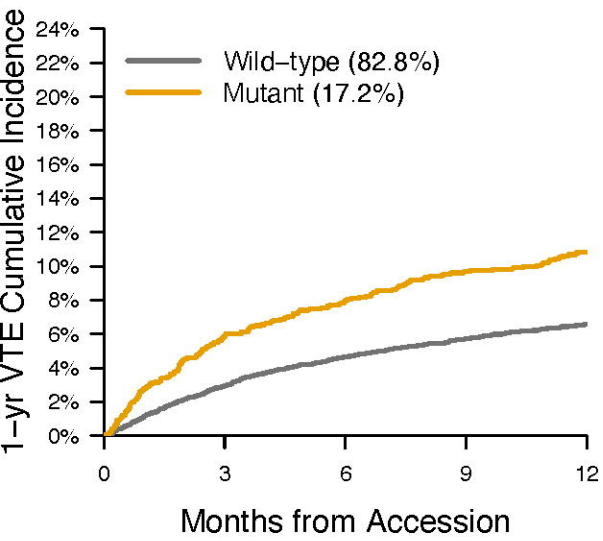
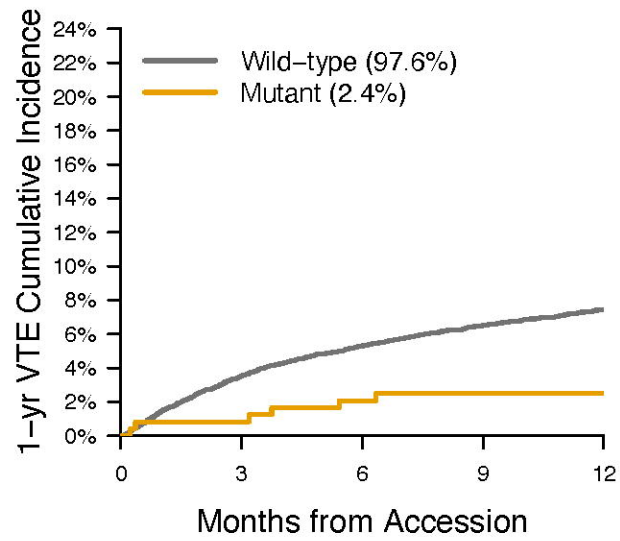
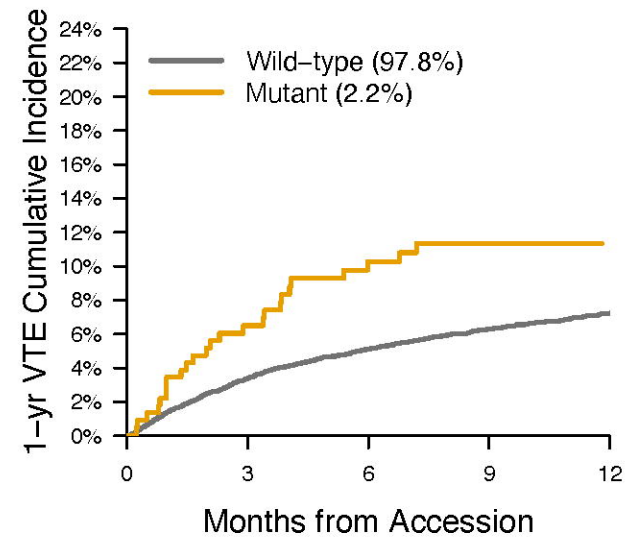
Individuals with a history of CAT after review by first observer: n=1,639

Individuals with a history of CAT confirmed by second observer: n=1,607

Additional patients found on NLP-driven search of clinical notes and radiology reports: n=355

Patients with at least one episode of CAT: n=1,962



**STK11****CDKN2B****KEAP1****KRAS****SETD2****CTNNB1****MET**