Genomic landscape and clinical utility of Korean advanced pan-cancer patients from prospective clinical sequencing: K-MASTER program

Kyong Hwa Park¹,², Jung Yoon Choi¹,², Ah-reum Lim¹,², Ju Won Kim¹,², Yoon Ji Choi¹,², Soohyeon Lee¹,², Jae Sook Sung², Hee-Joon Chung², Byunghyun Jang³, Dayoung Yoon³,⁴, Sukwon Kim³, Jason K. Sa³,*, Yeul Hong Kim¹,²,*

¹Division of Medical Oncology/Hematology, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea
²K-MASTER Cancer Precision Medicine Diagnosis and Treatment Enterprise, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea
³BK21 Graduate Program, Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Republic of Korea
⁴KU-KIST Graduate School of Converging Science and Technology, Korea University, Seoul, Republic of Korea

Running title: Genomic landscape of Korean advanced pan-cancer patients

Keywords: Pan-Cancer, East Asian, Precision Medicine, Next-Generation Sequencing, Personalized Treatment

Conflict of Interest:
The authors declare no conflict of interest.

Acknowledgments
This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI17C2206 to K.H. Park, Y.H. Kim), National...
Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2020R1F1A1076444 to J.K. Sa), Bio & Medical Technology Development Program of the NRF funded by the Korean government (MSIT) (2020M3A9D8038658 to J.K. Sa), Korea Evaluation Institute of Industrial Technology (KEIT) Grant funded by the Korean government (MSIT) (20009125 to J.K. Sa).

Correspondence to:
Jason K. Sa, PhD
Department of Biomedical Sciences, Korea University College of Medicine
73 Goryeodae-ro, Seongbuk-gu, Seoul, Republic of Korea
jasonksa@korea.ac.kr
+82-2-2286-1468

YeuIl Hong Kim, MD, PhD
Division of Medical Oncology/Hematology, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea
73 Goryeodae-ro, Seongbuk-gu, Seoul, Republic of Korea
yhk0215@korea.ac.kr
+82-2-920-5569
Abstract

The fundamental principle of precision oncology is centralized on identification of therapeutically exploitable targets that provides individual cancer patients an opportunity to make informed decisions on a personalized level. To facilitate and adopt such concepts within clinical practice, we have initiated a nation-wide, multi-institutional precision oncology screening program to examine and enroll patients into the most appropriate clinical trial based on their unique molecular properties. To determine prevalence of essential major driver mutations and to explore their dynamic associations at both single molecular and pathway levels, we first present a comprehensive overview on the genomic properties of East Asian cancer patients. We further delineated the extent of genomic diversity as well as clinical utility between patients from western and eastern cultures at the pan-cancer and single tumor entity levels. To support fellow oncology communities in future investigations involving large-scale analysis, all data have been made accessible to the public (https://kmportal.or.kr).

Statement of Significance:

We present a comprehensive overview on molecular properties of East Asian pan-cancer patients and demonstrate significant diversity in terms of genomic characteristics as well as clinical utilities compared to patients with European ancestry. The results of this study will lay the groundwork for designing personalized treatments in the clinical setting.
Introduction

The demand for personalized treatment has increased substantially among cancer patients with unmet clinical needs(1-5). In the oncology field, genomic biomarkers have been widely accepted as a new paradigm for patient treatment as well as drug discovery(6,7). Significance of molecular profiling via clinical next-generation sequencing (NGS) has been well demonstrated by precedent programs, including MSK-IMPACT and NCI-MATCH (8-14). Previous studies that have enrolled patients across a broad range of different tumor types, have made tremendous contributions in understanding the complexity of cancer genome at the pan-cancer level(15-17). However, as these studies have enrolled patients mainly of European origin, it has been challenging to directly implement such profound insights within clinical practice for East Asian cancer patients. A substantial number of studies have robustly demonstrated extensive genomic diversity as well as clinical utility among patients from distinct ethnic populations(18-22). Thus, it is of the utmost priority to establish a collection of cancer genomes, specializing on East Asian populations, across a wide spectrum of different cancer types.

Toward this goal, we have initiated the K-MASTER enterprise to collect and characterize the complex genomes of Korean patients with advanced solid tumors. We have leveraged previously established and validated clinical NGS panels to capture and detect major genomic aberrations, including single nucleotide variations (SNVs), small insertions and deletions, copy number alterations, and selected structural variations in cancer-related genes. Herein, we report the first phase of the K-MASTER precision oncology initiative, focusing on genomic characteristics of 4,028 pan-cancer patients to identify molecular signatures that constitute unique properties of cancer patients of East Asian ancestry.
Results

A schematic overflow of prospective clinical sequencing in Korean pan-cancer patients

Since the launch of the K-MASTER enterprise in June 2017, 4,028 Korean patients with advanced solid tumors have been subjected to prospective clinical sequencing (Supplementary Table 1). All of the patients enrolled into the program were either those whose standard-of-care options have already been exhausted without any other alternatives or will be exhausted in the nearest future. Genomic DNA samples that have been isolated from tumor tissue specimens underwent quality control prior to being subjected to deep-coverage sequencing in order to capture potential genomic aberrations, including SNVs, small indels, copy number alterations and selected structure variations (Supplementary Fig. 1). All sequencing data have been further processed and uploaded into the main database. Final reports have been reviewed by oncologists on a weekly basis.

Comprehensive genomic landscape of major cancer-driver mutations in Korean advanced pan-cancer patients

Our pan-cancer cohort at K-MASTER constitutes over 24 major cancer types across 4,028 Korean patients who were diagnosed with a wide spectrum of different solid tumors, primarily including colorectal, breast, gastric, ovarian, lung adenocarcinomas and sarcoma (Fig. 1A). Three oncology-based sequencing panels that encompass over full coding exons of commonly mutated cancer genes have been used in this study (409, 375, and 183 genes for KM, CancerSCAN, and FIRST panels, respectively). The average read depth for all clinical panels exceeded more than 650x to ensure that all essential genomic alterations, even at subclonal levels were covered. A total of 156,233 nonsynonymous mutations, including missense, nonsense, in-frame indels, frameshift, and splice-site, were detected. Notably, a subset of tumors demonstrated an excessive amount of tumor mutational burden, despite the limited coverage of targeted-exome sequencing. Through various machine-learning
algorithms(23,24), we discovered that several of these tumors harbored tracts of tandemly repeated DNA motifs and were thus classified as microsatellite instable (MSI) tumors. These MSI-high tumors were mainly detected in colorectal adenocarcinoma (COAD), gastric adenocarcinoma (STAD), and uterine corpus endometrial carcinoma (UCEC). Such results were consistent with previous studies that have reported on high prevalence of microsatellite instability in these cancerous tumors. Since these patients could potentially benefit from immune checkpoint blockades(25-27), our study supported the feasibility of routine clinical sequencing to identify potential benefits of immunotherapy (Fig. 1B).

The most frequently altered genes in the K-MASTER cohort included mutations in TP53 (48.1%), APC (21.8%), KRAS (17.7%), PIK3CA (16.3%), LRP1B (15.2%), ATM (11.8%), ARID1A (11.1%), and ATRX (10.8%) (Fig. 1B). Inactivating mutation in TP53 was the most prevalent genomic event in a number of tumors, including COAD, breast adenocarcinoma (BRCA), STAD, ovarian carcinoma (OV), lung adenocarcinoma (LUAD), and bladder urothelial carcinoma (BLCA). TP53 mutations were also largely observed within the previously identified “hot-spots,” including missense mutations in R175, R248, and R273 and nonsense mutations in R196 and R213(28,29) (Supplementary Fig. 2A). APC mutation was the second most frequently mutated gene and it was predominantly found in COAD, prostate adenocarcinoma (PRAD), and UCEC. Activating mutation in KRAS was the third leading genomic aberration, primarily observed in COAD, pancreatic adenocarcinoma (PAAD), cholangiocarcinoma (CHOL), ampullary carcinoma (AMPC), and UCEC. Consistent with previous large-scale genomic profile studies, KRASG12 and KRASG13 mutations were the most commonly altered events, accounting for over 66% of all KRAS-mutant tumors(30-32) (Supplementary Fig. 2B). Other major “hot-spot” mutations consisted of PIK3CAE542, PIK3CAH1047, BRAFV600E, ATRXE2246, and NF1S1100. Moreover, we identified a list of mutations that were highly enriched in specific tumor types respectively, such as BRAF mutations in skin cutaneous melanoma (SKCM) and thyroid carcinoma (THCA), KMT2D mutations in BLCA, and UCEC, EGFR mutations in LUAD and thymoma (THYM), PTEN mutations in UCEC and CDH1 mutations in STAD.
To explore and determine the etiology of each tumor lineage, we investigated on the mutational transition, context, and signature of each disease(33-35). Of the six classes of base substitution in each mutation type, all tumors demonstrated robust presence of C>T transition at CpG trinucleotides, which has been speculated to be a direct result of endogenous mutational process via spontaneous deamination of 5-methylcytosine (Fig. 1B). Since this phenomenon was strongly associated with the age of the patient at cancer diagnosis, we discovered that mutational “signature 1” was one of the predominant signatures that were observed across all tumor types. Other prominent mutational signatures that substantially occupied the overall mutational profiles of each tumor class were signatures that were associated with failure of DNA double-strand break-repair through homologous recombination and a defective DNA mismatch repair (MMR) system. As these features are the hallmarks of DNA repair mechanisms that ensure essential cell homeostasis(36), we suspected that a majority of the solid tumors in Korean patients may have largely propagated from MMR deficiency-derived tumorigenesis. We also identified tumor type-specific mutational signatures, including signature 4 (cigarette smoking) in LUAD, alkylating agent and ultraviolet light exposure-associated signature in SKCM, and aflatoxin-associated signature in LIHC.

Next, we have assessed significance of each genomic alteration that may directly affect the tumorigenic mechanism behind individual cancer type. Notably, we have identified previously known associations, including enrichments of APC, TP53, KRAS and SMAD4 mutations in COAD, PIK3CA and GATA3 mutations in BRCA, and KRAS, ARID1A, EGFR, and PTEN mutations in PAAD, STAD, LUAD, and UCEC, respectively (Fig. 1C; Supplementary Table 2). Conversely, we discovered that mutations in TP53 were relatively scarce in a number of different tumors, including sarcoma (SARC), SKCM, cervical cancer (CESC), and kidney renal clear cell carcinoma (KIRC), compared to other tumor types. KRAS mutations were significantly less frequent in head and neck cancer (HNSC) and BRCA, while mutations in APC were rare events in PAAD. Collectively, our results demonstrate the clinical utility of sequencing panels to provide comprehensive understanding of cancer etiology and the unique molecular properties of each individual patient.
Molecular interactions of major canonical oncogenic pathways in a pan-cancer model

Cancer is a disease of the genome with profound genomic aberrations and complex cellular signaling networks. Although each individual cancer lineage manifests different cellular architecture and hierarchy, dysregulation in the core oncogenic pathways is one of the primary contributing components underlying malignant transformation (15, 37). As such, we sought to determine whether molecular alterations at the cellular signaling pathway level were associated with different tumor origin. When we performed consensus hierarchical clustering of individual tumors based on the presence and absence of genomic alterations which correspond to major canonical pathways, we identified 7 distinct clusters with diverse genomic backgrounds (Fig. 2A and Supplementary Fig. 3). Cluster 1 (C1) mainly consisted of tumors harboring multiple genomic aberrations that affect a wide array of different pathways, including p53, Notch, RTK-RAS, PI3K, Wnt, and Hippo. Clusters 2, 3, 4, 5, and 7 were defined by enrichments of p53, RAS, Wnt, PI3K, and Notch signaling pathways, respectively. Notably, Cluster 6 (C6) was largely composed of tumors lacking any individual mutations that affect previously well-known canonical oncogenic pathways, suggesting potential alternative avenues for these tumors to exploit in terms of tumor propagation. Interestingly, each cluster composition was comprised of different tumor types, albeit both C4 and C5 demonstrated large collections of COAD and BRCA, respectively. C4, represented by dysregulation in the Wnt pathway was largely comprised of COAD and PRAD tumors, while C5, marked by genomic aberrations in the PI3K pathway was more predominantly composed of BRCA, BLCA, and CESC (Fig. 2B).

Genomic alteration at the pathway level provides a broader perspective on functional synergies within cancer malignancy and may also reflect potential therapeutic resistance or evasion mechanisms. To explore the dynamic associations within molecular pathways, we constructed a Bayesian network-based probabilistic model to identify significant co-occurrences and mutual exclusivity among major cancer driver pathways. Upon construction of multiple networks, we discovered significant co-occurrences of TGFβ, RAS, PI3K, Hippo, and cell cycle with RTK, which highlight potential synergistic activations of individual pathways (Fig. 2C). Both RAS and PI3K
pathways contained multiple pairs that were also co-enriched. Conversely, we also found several mutually exclusive pairs, notably in the MYC pathway with RTK, TGFb, and P53, and the NRF2 pathway with RTK, and cell cycle. Mutual exclusivity of the NRF2 pathway with cell cycle and RTK has also been previously reported in TCGA cohorts (15), further consolidating our findings. These associations suggest that activation of either pathway was sufficient for tumor propagation and alterations of the two may render tumor cells more susceptible for adverse situations. Our results collectively demonstrate dynamic interactions among major canonical pathways within tumor malignancy and functional interactions which suggest the use of combinational strategies in clinical settings.

Ethnic diversification of pan-cancer genome

Recent studies have highlighted existence of genomic diversity based on distinct racial or ethnic populations(18,20,38-40). Thus, assessment of genetic ancestry at the pan-cancer level may provide unprecedented insights into alternative therapeutic vulnerabilities. Identification of molecular pathways that are actively enriched in a specific ethnic subpopulation could potentially facilitate exploration of new, personalized treatment options. Toward this goal, we leveraged TCGA cohorts, which were mainly comprised of patients from European ancestry, across 20 major cancer types to characterize and compare prevalence of major-cancer driver genomic aberrations at both single genomic and pathway levels. While several major canonical pathways, including TGFb, cell cycle, Myc, and Hippo constituted highly consistent levels of pathway dysregulation at the pan-cancer level between TCGA and K-MASTER, some oncogenic pathways demonstrated significant levels of genetic diversity (Supplementary Figs. 4-5). Surprisingly, while patients from K-MASTER showed significant levels of KRAS mutations, TCGA patients were characterized by recurrent mutations in BRAF. Ethnic-driven genomic diversity became increasingly more apparent when compared at individual cancer lineage levels. While the majority of the genomic alterations demonstrated considerable levels of similarity (r = 0.618, p < 2.2 x 10^{-22}), mutations in TP53 differed significantly in several pathological entities (Fig. 3A). For example, patients who were enrolled in TCGA and were diagnosed with either OV, esophageal
carcinoma (ESCA), HNSC, PAAD, or SARC demonstrated higher levels of TP53 ablation, whereas enrichment of TP53 mutation was more evident in K-MASTER COAD, BLCA, BRCA, CHOL, and PRAD patients. Furthermore, we discovered that TCGA COAD patients were marked by enrichment of somatic APC and PIK3CA mutations, while K-MASTER CESC patients showed frequent alterations in PIK3CA and ATRX. Additionally, SKCM patients from TCGA showed recurrent dysregulations in major oncogenic drivers, including BRAF, LRP1B, and FAT3.

Next, we analyzed all essential mutations that differed significantly between TCGA and K-MASTER at both individual cancer lineage and pan-cancer levels. As a result, we discovered 25 recurrently mutated genes (> 200 tumors) that demonstrated significant statistical differences. Among them, APC was the most significantly mutated gene in K-MASTER, followed by AR, KRAS, TP53, ATRX, MAP2K7, and ATM (Fig. 3B). Conversely, TCGA patients demonstrated predominance of BRAF, FAT1, PTEN, GATA3, EPHB1, AKT2, EPHA2, and EGFR mutations in both pan-cancer and individual tumor types. Furthermore, we discovered that mutations in MMR encoding molecules, including MSH3, MSH6, MLH1, and MSH2 were significantly more frequent in K-MASTER patients, consolidating previous speculations on dominance of MMR-deficiency associated mutational signature in our cohort. Furthermore, we have discovered a significant enrichment of IDH1 mutations only in TCGA cholangiocarcinoma patients, whereas patients from K-MASTER mostly lacked such genomic aberrations; rather they exhibited highly recurrent mutations in TP53 and KRAS (Fig. 3B and Supplementary Fig. 6). Interestingly, one of the earlier studies that took advantage of sequencing technology has identified recurrent mutations in multiple chromatin-remodeling genes, including IDH1 in intrahepatic CHOL(41). However, a majority of the enrolled patients were of European ancestry (93.3%). To further investigate upon the genomic diversity of CHOL at the ethnicity level, we curated additional mutational profiles of 195 CHOL patients from the Memorial Sloan Kettering (MSK)-IMPACT cohort(42) and 103 patients from the Eastern Hepatobiliary Surgery Hospital (EHSH), China(43). Remarkably, all key chromatin-remodeling genes, including IDH1, BAP1, and PBRM1 were significantly enriched only in the MSK-IMPACT cohort. Such results further advocated our previous findings (Fig 3C). Taken together,
our results demonstrate the significance of genomic diversity at both the pan-cancer and individual cancer lineage levels between populations from European and East Asian ancestries.

**Ethnic diversification of clinical actionability**

The molecular characterization of tumors enable patient-tailored therapy. Furthermore, a substantial number of studies have demonstrated remarkable clinical success through employment of essential molecular biomarkers, including trastuzumab against HER2 in BRCA(2), vemurafenib against $BRAF^{V600E}$ in melanoma(44), gefitinib against $EGFR$ in non-small cell lung cancer(4), and etc. Therefore, numerous institutions have begun to employ routine clinical sequencing to make informed decisions in the choice of targeted therapy. In hopes of identifying a subset of patients that may benefit from such therapeutic intervention, we evaluated clinically relevant molecular properties of the K-MASTER cohort compared to TCGA based on different disease backgrounds. Overall, THCA, UCEC, BRCA, and SKCM demonstrated the highest proportion of druggable mutations for both cohorts, while tumors with limited levels of targetable alterations were comprised of SARC, KIRC, and liver hepatocellular carcinoma (LIHC). These results urge the exploration of alternative therapeutic avenues (Fig. 4A). Compared to 31.8% of TCGA patients, 28.7% of K-MASTER patients possessed at least one or more molecular targets that were druggable (Supplementary Fig. 7). At the individual molecular level, somatic mutations in $KRAS$ and $PIK3CA$ provided the most therapeutic opportunities for COAD and BRCA patients, respectively (Supplementary Fig. 8). $BRAF$ mutation was the third leading target, enriched in COAD, CHOL, SKCM, and THCA. Other prospective targets consisted of $CDKN2A$ in HNSC, $ERBB2$ in STAD, $NRAS$ in SKCM, $PTEN$ in UCEC, $AKT1$ in BRCA, $EGFR$ in LUAD, $BRCA1/2$ in OV, $FGFR3$ and $ERCC2$ in BLCA, and $KIT$ in SARC.

Next, we leveraged the OncoKB knowledge database to systematically annotate each molecular alteration based on clinically utility(45). Among 963 pharmacogenomic associations that were previously reported, $BRAF$ V600E mutation was the most prevalent genomic aberration, primarily observed in COAD, SKCM and THCA, followed by $KRAS$ G12D and $PIK3CA$ E545K and H1047R mutations (Fig. 4A and
Supplementary Fig. 9A). Strikingly, the level at which \textit{BRAF} V600E mutation occurred significantly differed between TCGA and K-MASTER, where the patients from TCGA harbored the alteration at a significantly higher rate (73.3% vs. 25.7%; \textit{P-value} = 1.58x10^{-36}) (Supplementary Fig. 9B). While the level of treatment opportunity in the K-MASTER and TCGA cohorts was similar among several major solid tumors (COAD, BRCA, BLCA, PAAD, UCEC and THCA), other cancer types manifested significant differences. STAD patients from TCGA were characterized by recurrent mutations in \textit{ERBB2} and \textit{PIK3CA}, while K-MASTER patients showed higher frequency of \textit{KRAS} and \textit{PIK3CA} mutations in OV. Moreover, we discovered that both PRAD and KIRC patients from K-MASTER possessed several clinically targetable aberrations, including \textit{EGFR}, \textit{KRAS}, and \textit{PIK3CA}, whereas patients from TCGA completely lacked such opportunities. In addition, TCGA LUAD patients exhibited higher activation of various \textit{KRAS} mutations, whereas K-MASTER patients were marked by enrichments in \textit{EGFR} exon 19 deletion, which encodes part of the kinase domain and renders higher susceptibility of tumors to various EGFR tyrosine kinase inhibitors, including gefitinib and afatinib.

Patients were enrolled into 20 distinct matched clinical trials depending on the unique molecular and histopathological properties of each patient. Through such assignment, we were able to target various major molecular aberrations, ranging from receptor tyrosine kinases, including EGFR, c-MET, and HER2 to immune checkpoint molecules, such as PD-1 and PD-L1 (Fig. 4B). A total of 440 patients have been enrolled on the basis of corresponding molecular targets so far. The most common treatments consisted of EGFR, HER2, PIK3CA, PD-1 and PD-L1 inhibitors, including gefitinib and lazertinib in non-small-cell lung carcinoma, varlitinib in STAD, trastuzumab-biosimilar in BRCA, STAD, and BLCA, gedatolisib in BRCA, alpelisib in COAD and other solid tumors, nivolumab in solid tumors with HRD/dMMR deficiency, PDR001 in ESCA, pembrolizumab in NSCLC, avelumab in MSI-H/\textit{POLE}-mutant COAD, and INCMGA00012 in ESCA. Notably, a BRCA patient with liver metastasis, harboring activation of HER2 and \textit{PIK3CA} H1047R mutation, demonstrated exceptional response to third-line treatment of gedatolisib with trastuzumab (Fig. 4C). The patient was previously treated with docetaxel, trastuzumab, and pertuzumab as first-line therapy and T-DM1 as second-line treatment. However, the patient eventually experienced
tumor relapse without any alternative options. Based on the patient’s unique molecular property, a combinational treatment of PI3K inhibitor with trastuzumab was given and the patient has shown remarkable clinical response within the first six weeks and the response prolonged even after 24 weeks. Collectively, our results demonstrate considerable differences in terms of clinical applicability between East Asian and European American cancer patients and recommend that the application of molecular-guided treatments should be implemented in caution depending on the ethnic background of the patient population.

Discussion
In the present study, we report on the clinical utility and significance of systematic prospective sequencing to delineate and identify the major genomic aberrations in East Asian pan-cancer patients. This study which involved nation-wide, multi-institutional effort demonstrated clinical feasibility in guiding individual patients to ideally matched clinical trials that may provide the maximum therapeutic efficacy over a wide spectrum of different cancer types. Based on this enterprise, we presented the first phase of our precision oncology initiative, focusing on a comprehensive molecular landscape in 4,028 East Asian pan-cancer patients. By utilizing an integrative analytical approach, we have identified key driver alterations and etiologies of 24 major cancer lineages, and dynamic interactions at both single molecular and pathway levels. Furthermore, to investigate upon the impact of ethnic ancestry on both genomic diversity and clinical utility, we leveraged previously established large-scale pan-cancer genomic cohorts, including TCGA and MSK-IMPACT. Our results provided a comprehensive overview on therapeutic opportunities for individual cancer types based on distinct ancestral populations. Notably, we have identified considerable levels of differences between eastern and western populations in terms of both individual genomic events and pathway levels. Activations of NRF2 and PI3K signals were predominantly observed in TCGA cohort, while K-MASTER patients demonstrated recurrent ablations in RTK-RAS, TP53, WNT, and Notch encoding molecules. This notable distinction was consistently observed in previous studies where patients of European American (EA) ancestry showed higher rates of dysregulation in the PI3K encoding genes, including PTEN.
compared to African American (AA) populations (18). While AA patients were marked by enrichments of p53 pathway abnormality, specifically through genomic alteration in TP53, our study showed that East Asian patients were constituted by activation of the RTK-RAS and Notch pathways. Furthermore, we have discovered higher dysregulation in the MMR encoding molecules for K-MASTER patients, which further consolidated previous observations on enrichment of MMRd-associated mutational signatures. These results collectively suggest that Korean cancer patients may have largely propagated from impairment of DNA repair mechanisms. Such genomic disparity was further evident when assessed at individual cancer levels. Recurrent mutations of chromatic-remodeling genes such as BAP1, PBRM1 and IDH1 were relatively more scare in cholangiocarcinoma patients from Eastern backgrounds. Furthermore, we have discovered significances differences in molecular-based treatment opportunities at both pan-cancer and individual tumor levels. Notably, we have discovered that BRAF V600E mutation was significantly more enriched in TCGA patients, while K-MASTER patients harbored considerable levels of KRAS G12 mutations. Other key findings included predominance of mutations in NRAS at both Q61R and Q61K domains only in TCGA and substantial differences of clinical actionability profiles in STAD, OV, CHOL, PRAD, KIRC, LIHC, and ACC between TCGA and K-MASTER cohorts. On the contrary, we also found that COAD, BRCA, BLCA, PAAD, UCEC and THCA tumors shared similar molecular properties. While our results have demonstrated profound levels of genomic diversity between patients from different ethnic origins at the pan-cancer level, there were several limitations to this approach. First, the TCGA cohort was mainly comprised of tumors that were derived from a single distinct pathological entity while the K-MASTER cohort originated from more generalized classifications. Furthermore, there were also discrepancies in terms of the cohort size at the individual disease level. However, since to the best of our knowledge, as this was the first comprehensive molecular study to evaluate genomic profiles of pan-cancer patients in the East Asian population, our results highlight the significance of employing ethnic-based personalized approach in cancer therapy. Moreover, the current study provides a more appropriate real-world scenario of cancer treatment where the patients have been diagnosed with an
advanced stage of tumor and are searching for alternative therapeutic avenue; while other large-scale genomic studies, including TCGA focused primarily on untreated tumors at a relatively early stage in tumor progression. As such, we speculate that our study provides higher clinical applicability. Furthermore, the primary utility of the K-MASTER program lies in its ability to provide sufficient evidence to make informed decisions in terms of patient treatment. Based on these evidence, 440 patients have been enrolled into 20 different clinical trials targeting major genomic aberrations across a broad range of different tumor types. At this point, a majority of the clinical trials are still ongoing, and thus we were not able to disclose the full results in the present study. However, as a proof-of-concept, we showed that a BRCA diagnosed patient who harbored *PIK3CA* activating mutation with liver metastasis, demonstrated a remarkable response to PIK3CA-mediated therapy through prospective clinical sequencing.

Both clinical feasibility and adaptation of precision oncology are determined by several key components that need to be recognized and accounted for. The main priorities of patients are, accuracy and timeliness of the results, access to the treatment, cost-benefit ratio, and the degree of improvement in quality of life through new treatments. For clinicians, on the other hand, turn-around time, accuracy of genomic screening tests, reimbursement rate, and possibility of evidence-based clinical decision are the primary concerns. Through the K-MASTER enterprise, 55 cancer-treating hospitals and centers have participated and experienced a full cycle of personalized care (from NGS testing to making clinically informed decisions) within a short period of time. We strongly believe that the K-MASTER initiative clearly exhibits how quickly and efficiently a public-initiated program can expand the potential of precision oncology, even for latecomers. In the next phase of our study, we aim to explore dynamic interactions of genomic aberrations with therapeutic responses from 20 different clinical trials at the pan-cancer level.
Methods

K-MASTER initiative and cancer specimens collection

K-MASTER initiative is a government-supported precision medicine enterprise that mainly focuses on diagnosis and treatment of cancer patients (https://k-master.org/eng.php). The primary objective of the operation is to collect and characterize the complex genome of 10,000 Korean patients with advanced solid tumors who have been enrolled in the master screening protocol, KM-00. Based on patients who were initially screened using the KM-00 protocol and have been identified with at least one actionable therapeutic target of treatment, K-MASTER has initiated 20 distinct clinical trials using single or combination targeting agents. Patients with advanced solid tumors were enrolled to master screening protocol of KM-00 at one of the 55 participating sites after IRB approval. After receiving written informed consent from the patients, tumor tissue specimens were collected and archived. Both clinical and genomic information have been stored in the K-MASTER database and has been released to the public (https://kmportal.or.kr). As of December 2020, more than 7,900 patients have participated and were enrolled in the KM-00 master screening program. Furthermore, we have developed a “Match Master System” that utilizes OncoKB knowledge database for clinical decision support on actionability of the genes. We have also leveraged updated results from all relevant clinical trials that have been conducted. The research conformed to the principles of the Helsinki Declaration.
K-MASTER panel sequencing

K-MASTER employed two previously established tissue-based NGS panels (FIRST and CancerSCAN) to detect major genomic aberrations, including mutations, copy number alterations (CNAs), and small insertions and deletions in cancer-related genes. CancerSCAN has been further upgraded to K-MASTER v1.0 and v1.1. Genomic DNA from formalin-fixed paraffin embedded (FFPE) samples or plasma were extracted using the QIAamp FFPE Tissue kit (Qiagen) or QIAamp circulating nucleic acid kit (Qiagen), respectively. cfDNA purity was measured using an Agilent High Sensitivity DNA Kit and 2100 bioanalyzer instrument (Agilent Technologies). When required, additional purification was performed using Agencourt AMPure XP (Beckman Coulter) to further remove contaminating nucleic acid. Centrally isolated genomic DNA samples that underwent QC were sent to the K-MASTER genomic analysis laboratories.

Mutation calls

The sequenced reads from the FASTQ files were aligned to the human genome assembly (hg19) using Burrows-Wheeler Aligner. The initial aligned BAM files were further subjected to pre-processing steps, including sorting, removal of duplicated reads, local re-alignment around small indels and recalibration of base quality scores using SAMtools, Picard, and Genome Analysis Toolkit (GATK). To make high-confidence predictions on mutation calls, we used MuTect2. 1000 Genome, gnomAD, and dbSNP datasets were used as reference database for known polymorphic sites. We used Variant Effect Predictor (VEP) to annotate each variant. Mutations with a minimum depth ≥ 20 and variant allele frequency of ≥ 2 were used in this study.

Mutational signatures

We used deconstructSigs in R to perform mutational signatures analysis. It uses a list of mutations based on six substitution classes (C>T, C>A, C>G, T>C, T>A, T>G) and base contexts immediately before and after the mutated nucleotide within the exome regions. It also generates a composition of a given set of mutational signatures that were previously identified. Thirty different mutational signatures from “signature.cosmic” was used as a reference signature and were represented in the following terms: Age
(Signature 1), APOBEC (Signature 2), DNA-DSB (Double Strand Break; Signature 3), Smoking (Signature 4), Sig 5 (Signature 5), MMRd (MisMatch Repair-deficiency; Signature 6), UV (Signature 7), Alkylating (Signature 11), APOBEC2 (Signature 13), Aflatoxin (Signature 24), Tobacco (Signature 25), and Sig 30 (Signature 30). To prevent overestimation of mutational signature proportion from patients with a limited number of mutations, we only selected tumors with more than 20 nonsynonymous mutations. We filtered mutation signatures that were present in at least 5% of the samples in each tumor type.

Microsatellite instability status

MSIseq in R was used to assess tumor microsatellite instability. MSIseq is a decision tree classifier using a list of mutations based on different mutation rates in all sites as well as in simple sequence repeats. It uses previously established somatic mutation data from the exomes of 361 tumors as a training set and classifies newly generated tumors based on annotation of locations of simple-sequence repeats and sequence length of each tumor.

Comparison of genomic diversity and pathway activity between K-MASTER and The Cancer Genome Atlas (TCGA) cohorts

To compare the frequencies of major genomic aberrations based on ethnicity, we acquired TCGA pan-cancer somatic mutation data and clinical data resource from Genomic Data Commons. Only mutations that were annotated as “PASS” in the “FILTER” column have been retained for all cancer types except for earlier TCGA samples which were sequenced using the whole genome amplified method. Afterwards, we only selected mutation data for patients who were diagnosed with the matching tumor types characterized in the K-MASTER program, including colon adenocarcinoma, breast invasive carcinoma, stomach adenocarcinoma, ovarian serous cystadenocarcinoma, head and neck squamous cell carcinoma, lung adenocarcinoma, cholangiocarcinoma, bladder urothelial carcinoma, pancreatic adenocarcinoma, skin cutaneous melanoma, prostate adenocarcinoma, kidney renal clear cell carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, uterine corpus
endometrial carcinoma, esophageal carcinoma, liver hepatocellular carcinoma, adrenocortical carcinoma, thymoma, and thyroid carcinoma. This resulted in a total of 7,557 patients. Next, we selected for patients who have been annotated as “WHITE” in the “race” column, resulting in the final list of 5,579 patients. Only genes that were captured from the K-MASTER sequencing panels underwent further selection. Among them, only the protein-coding mutations were considered for comparison analysis. For the pan-cancer comparative analysis, we leveraged mutation profiles of the resulting TCGA patients with the K-MASTER cohort and compared the frequency of each individual mutation corresponding to major oncogenic canonical pathway. Pathway diagrams and genes were curated from PathwayMapper, a tool that provides visualization and design of major oncogenic pathways from previous TCGA publications. This tool is publicly available online at www.pathwaymapper.org. To perform individual tumor-type comparison, we curated mutation profiles from both TCGA and K-MASTER cohort based on individual cancer lineage and compared the overall mutational frequency. For Cholangiocarcinoma patients, we also acquired mutation profiles for MSK-IMPACT and EHSH cohorts from the cBioPortal (http://www.cbioportal.org). Chi-square tests were performed to compare the frequencies of individual gene-level mutations between K-MASTER and TCGA cohorts and corrections for multiple-hypothesis testing were performed to account for false discovery rate.

Clinical trial enrollment
Genetic alterations, including single nucleotide variants, insertions, deletions (indels), copy number alterations, or structural rearrangements with clinical actionability were reported in a clinical report format. Treatment options including clinical trials in the K-MASTER program were recommended based on the OncoKB knowledge database (OncoKB API;2020.12) and inclusion criteria for each trial.

Statistical analyses
All statistical analyses were performed using the R software 3.4.0. (https://www.r-project.org).
Data availability

All data have been deposited and hosted on our portal at https://kmportal.or.kr.

References


Figure Legends

Figure 1. Genomic landscape of K-MASTER pan-cancer cohort. A) Distribution of major tumor types and corresponding clinical sequencing panels from 4,028 pan-cancer patients. B) Genomic landscape of major cancer driver mutations based on distinct tumor types. The top panel depicts number of nonsynonymous mutations with microsatellite instability status. The top middle panel shows frequency of each mutation in corresponding tumor types. Red indicates activating oncogenes and blue indicates inactivating tumor suppressors. The left bar indicates the percentage of tumor within the pan-cancer cohort in respect to different type of mutation. The bottom middle panel exhibits six classes of base substitution in each mutation type. The bottom panel represents mutational signatures. The size of the node is proportional to the number of patients within each tumor type. C) Volcano plot representation of tumor frequency differences (x-axis) between tumors with the mutation in the corresponding tumor type vs. rest and its significance (y-axis). Mutations that are significantly more enriched in specific tumor types are in colored in red while absences are colored in blue.

Figure 2. Major oncogenic canonical pathways of K-MASTER cohort. A) Unsupervised hierarchical clustering of pan-cancer patients based on major oncogenic canonical pathways. Patients have been marked with a mutation for each pathway if the
patient harbors at least one mutation that belongs to the corresponding pathway. B) Pie chart distribution of patients within each corresponding cluster (C4: Top; C5: Bottom). The percentage represents the frequency of patients that belong to the respective pathway cluster within the corresponding tumor type (i.e. 23.9% of COAD patients and 13.6% of PRAD patients belong to C4 cluster). C) Bayesian network analysis depicting the co-occurrences and mutual exclusivity of major canonical pathways in the K-MASTER pan-cancer cohort. Only the significant associations are shown.

**Figure 3. Genomic diversity of pan-cancer patients based on ethnicity.** A) Mutation frequencies in TCGA (y-axis) and K-MASTER cohorts (x-axis). The color corresponds to distinct tumor type and size of each node represents the number of tumors within K-MASTER cohort. B) Pan-cancer meta-analysis on recurrent mutations between K-MASTER and TCGA cohorts across 20 cancer types. The top bar graph demonstrates the significance of each mutation at the pan-cancer level. The bottom dot plot represents the significance and difference of each mutation at individual cancer lineage levels. The color corresponds to the effect size of K-MASTER patients compared to that of the TCGA and the size is proportional to the significance. C) Ternary diagram depicting mutation frequencies in K-MASTER, Eastern Hepatobiliary Surgery Hospital (EHSH) in Shanghai, China and Memorial Sloan Kettering (MSK)-IMPACT cohort. The size of each node represents the number of tumors with respect to the mutation in the K-MASTER cohort and the color spectrum indicates the significance of relative frequencies.

**Figure 4. Therapeutical landscape of pan-cancer patients based on ethnicity** A) The top bar graph depicts the distribution of major mutations based on the clinical actionability of the mutations in the K-MASTER cohort. The bottom bar graph represents the TCGA pan-cancer cohort. Mutations are categorized as “druggable” if they are FDA-approved biomarkers for FDA-approved drugs. Mutations are labeled as “actionable” if there is substantially compelling evidence to support the use of biomarkers to predict the response, especially the resistance of FDA-approved drugs. Mutations are labeled as “COSMIC” if they have been previously annotated using the
COSMIC database. Mutations that do not belong in any of the categories mentioned previously are depicted as “SNV.” The middle panel illustrates the frequencies of clinically actionable mutations across major cancer types in K-MASTER (left portion of the cell) compared to TCGA (right portion of the cell). Genes have been grouped by pathway. The white asterisk represents tumor types with higher than 25% frequency (29.1% of KRAS G12D mutation in K-MASTER PAAD, 43.6% of BRAF V600E mutation in TCGA SKCM, and 61.5% and 59.3% of BRAF V600E mutations in K-MASTER and TCGA THCA, respectively). B) The number of patients who have been enrolled to matched clinical trials based on unique molecular alterations. “Trastuzumab bs” indicates Trastuzumab biosimilar. C) Clinical course of third line trastuzumab and gedatolisib treatment in a BRCA patient with liver metastasis, harboring PIK3CA H1047R mutation. T1-weighted contrast enhanced magnetic resonance images are shown for baseline, 6-week, and 24-week post-treatment. Orange arrows indicate measurable tumors.
Co-occurrence
Mutually exclusive
q-value < 0.05

RTK
Cellcycle
NOTCH
NRF2
PI3K
P53
MYC
WNT
TGFB
RAS
HIPPO

Figure 2

23.9%
C4 Cluster
13.6%
14.5%
C5 Cluster
8.8%
14.9%

Research.
EGFR 746 750del

PIK3CA H1047R

CDKN2A W110*

PIK3CA H1047L

CDKN2A P114L

PIK3CA M1043I

PIK3CA Q546R

PIK3CA R108H

PIK3CA Q546K

MAP2K1 P124S

PIK3CA C420R

PIK3CA N345K

CDKN2A H83Y

PIK3CA E545K

PIK3CA E545A

PIK3CA E542K

PIK3CA E453K

FGFR2 S252W

ERBB2 R678Q

FGFR3 S249C

CDKN2A R80* 100%

CDKN2A R58* 60%

CDKN2A P114L 80%

CDKN2A H83Y 100%

ERBB2 L755S 20%

PIK3CA R93Q

PIK3CA R88Q

SF3B1 K700E

ERBB2 V842I

PTEN R130Q

PIK3CA E545K

PIK3CA E545A

PIK3CA E542K

PIK3CA E453K

ERBB2 V842I
Genomic landscape and clinical utility of Korean advanced pan-cancer patients from prospective clinical sequencing: K-MASTER program

Kyong Hwa Park, Jung Yoon Choi, Ah-reum Lim, et al.

Cancer Discov  Published OnlineFirst December 3, 2021.