이학박사 학위논문

Analysis of the relationship between hidden inherited germline factor and cancer risk

숨겨진 유전성 생식세포 돌연변이와 암발생 위험성의 상관관계 분석 연구

2019 년 08 월

서울대학교 의과대학

김 대 윤

A thesis for the PhD Degree

Analysis of the relationship between hidden inherited germline factor and cancer risk

숨겨진 유전성 생식세포 돌연변이와 암발생 위험성의 상관관계 분석 연구

August 2019

Seoul National University

College of Medicine

Daeyoon Kim

숨겨진 유전성 생식세포 돌연변이와 암발생 위험성의 상관관계 분석 연구

지도 교수 윤 성 수

이 논문을 이학박사 학위논문으로 제출함
2019년 4월

서울대학교 대학원
의과대학 협동과정 종양생물학 전공
김 대 윤

김 대 윤의 이학박사 학위논문을 인준함
2019년 7월

위 원 장 _____ (인)
부위원장 _____ (인)
위 원 _____ (인)
위 원 _____ (인)
위 원 _____ (인)

Analysis of the relationship between hidden inherited germline factor and cancer risk

by
Daeyoon Kim
(Directed by Sung-Soo Yoon, M.D., Ph.D.)

A thesis submitted to the Department of Medicine in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at Seoul National University College of Medicine

July 2018

Approved by Thesis Committee:

Professor _____ **Chairman**

Professor _____ **Vice Chairman**

Professor _____

Professor _____

Professor _____

ABSTRACT

Introduction: Previous cancer-associated germline mutation studies have highlighted tumor suppressor genes such as *TP53, APC, BRCA2, NF1, PMS2,* and *RB1*. Many studies on tumor suppressor genes have been conducted because of the high penetration rate of pathogenic mutations. For this reason, cancer-related germline studies have been focused on tumor suppressor genes, and the role of germline genes other than tumor suppressor genes remains unclear. However, the possibility of presenting non-tumor suppressor genes that affect tumor development is not negligible. The mechanism of tumorigenesis by these non-tumor suppressor genes that exist as germline mutations contribute differently to tumor development. In this study, we will discuss the relationship between two different types of non-tumor suppressor genes and tumorigenesis. The first is the germline mutation of oncogenes in Acute Myeloid Leukemia(AML) patients, and the second is the correlation between primary immunodeficiency-associated gene mutations and tumor risk.

Methods: To determine the oncogenic germline mutation of AML, WES data of 76 individuals from Seoul National University were analyzed. The correlation between immunodeficiency disease and cancer development was assessed using ICGC WGS data from 2566 individuals. For the normal control, WGS data of 2504 individuals were obtained from the 1000 Genome Project, and tumor susceptibility was examined by comparing the cancer and control groups.

Results: We focused on the germline *FLT3* as an oncogenic gene that increases the incidence of AML. An *FLT3* germline mutation in four loci was identified in 76 patients with AML, and the cancer incidence was confirmed to increase in the locus corresponding to D358V. The overall survival of patients with D358V was significantly decreased, indicating that the mutation had a negative effect on the prognosis of patients.

Regarding the association between PID and cancer, it has been confirmed that the PID-related

genes increase cancer risk by 1.5 fold. Seven carcinomas were also associated with increased cancer risk, and the age of onset was significantly reduced in hepatocellular carcinoma and bladder cancer patients with PID.

Conclusions: We revealed the association between two germline mutations and increased cancer risk, which was previously undefined owing to the rare frequency. The first germline mutation gene was the *FLT3* mutation known as the oncogene. *FLT3*, which exists as a germline mutation, is presumed to contribute to the development of tumor cells by reducing *FLT3* activity and turning on alternative cell signal transduction, unlike the *FLT3* somatic mutation that induces cell division of tumor cells. The second germline mutation is a primary immunodeficiency disease-associated gene. Loss-of-function in immunodeficiency-associated genes does not directly induce cancer development but is assumed to promote tumorigenesis owing to a diminished immune surveillance effect. In addition to germline factors identified in this study, more novel germline factors are expected to be revealed. These efforts will contribute to the development of prophylactic treatments to prevent cancer patients from experiencing a relapse after treatment.

Keywords: Germline Mutation, Inheritance Cancer, Next Generation Sequencing
Student number: 2016-30613

CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

Chapter 2

GENERAL INTRODUCTION

The “two-hit” hypothesis of Alfred Knudson began to identify cancer predisposition genes (CPGs), which are associated with an increased risk of cancer germline mutations (Rahman, 2014). Through a statistical analysis of retinoblastoma cases, Knudson proposed that “two hits” to the DNA were necessary to cause cancer. The “first hit” is an inherited variation in one allele of the gene while the “second hit” is a somatically acquired inactivation of the second allele (Knudson, 1971). This model was confirmed by the identification of biallelic inactivation of the *RB1* gene in retinoblastoma, and most known high-penetrance inherited cancer predisposition variants are loss-of-function mutations in recessively acting tumor suppressor (TS) genes (Knudson, 2001; Rahman, 2014).

Additionally, previous pan-cancer studies have highlighted pathogenic germline variants in tumor suppressor genes, including *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, and *PALB2* in adult cancer cases recorded in The Cancer Genome Atlas (TCGA) (Lu et al., 2015) and the Collaborative Oncological Gene-Environment Study (COGS) (Southey et al., 2016), as well as *TP53*, *APC*, *BRCA2*, *NF1*, *PMS2*, and *RB1* using 1,120 pediatric cancer cases from the Pediatric Cancer Genome Project (PCGP) (Zhang et al., 2016).

Along with this historical background, the study of the interaction between genes, other than tumor suppressor genes, and cancer risk has not been investigated. In this paper, we discuss the relationship between two types of

germline mutations and cancer risk in addition to tumor suppressor genes, which have gone unnoticed in the past.

The first type of germline mutation is the oncogenes mutation in cancer patients. Despite a close correlation between gene function and cancer development, epidemiologic studies have not recognized the oncogene as a germline mutation that increases cancer risk. However, *HRAS*, *EGFR*, and *KIT* genes that increase tumor susceptibility to rhabdomyosarcoma, ganglion cell tumor, and bladder cancer were reported as oncogene-predisposing germline mutations (Aoki et al., 2005; Centeno et al., 2011; Hirota et al., 2002). This phenomenon supports the assumption that other oncogenic genes may be present.

The second type of germline mutation is the development of tumors due to the haploinsufficiency phenomenon, especially the reduction of immune surveillance as a result of the dysfunction of primary immunodeficiency disease-associated genes. In the case of primary immunodeficiency disease (PID), the direct association between PID and cancer development was not known, so it was not considered as a germline factor. However, the association between various severe PIDs and cancer was revealed through epidemiological analysis of the multicenter registry of patients with PID (Kersey et al., 1974; Mayor et al., 2018; Rodrigues et al., 2004).

These germline factors are evolutionarily at a very low frequency in the general population even though they are relevant to the disease, so it is

difficult to determine their relationship with cancer development. However, the past decade has seen an exponential increase in the number of genomic data, enabling researchers to discover germline factors that were previously hidden. Based on this trend, we made an effort to reveal novel germline factors and will present our results in this paper as follows: 1) Germline mutations of *FLT3* that are associated with increased cancer risk and poor prognosis in acute myeloid leukemia (AML) patients; 2) Primary immunodeficiency disease and cancer risk using the Korean AML cohort and International Cancer Genome Consortium (ICGC), respectively.

CHAPTER 1

Recurrent germline FLT3 pathogenic variant (D358V) is associated with a high prevalence of hematopoietic malignancy and poor prognosis in acute myeloid leukemia

INTRODUCTION

The idea that cancer is formed through the accumulation of multiple genetic alterations is known as the two-hit or multiple-hit hypothesis. Carl O. Nordling first proposed the multiple-hit hypothesis in 1953 (Nordling, 1953) when he suggested that multiple hits to DNA were necessary to cause cancer. Genetic alternation can be a combination of multiple somatic pathogenic variants or a combination of somatic and germline pathogenic variants. Traditionally, somatic pathogenic variants have received the majority of research focus, and few studies have investigated the role of germline pathogenic variants.

Acute myeloid leukemia (AML) is a form of cancer characterized by infiltration of the bone marrow, blood, and other tissues by proliferative, clonal, and abnormally differentiated cells of the hematopoietic system (Dohner et al., 2015). Conventionally, the French-American-British (FAB) classification has been used to classify AML. The FAB classification is based on cell morphology and not disease mechanism; therefore, it is difficult to predict prognosis and select optimal therapeutic approaches for AML using this classification system. As a result, pathogenic variants in several genes, including *FLT3*, *RUNX1*, *PMA-RARA*, and *NPM1*, which have been repeatedly observed in AML, are used in disease classification. Of these pathogenic variants, those affecting *FLT3* (FMS-Like Tyrosine Kinase 3, also known as *FLK2* [fetal liver kinase 2] and *STK1* [human stem cell tyrosine kinase 1]), which plays a key role in hematopoiesis and leukemogenesis are some of the most important in AML. Although many AML patients achieve complete remission (CR) with conventional induction chemotherapy, a significant proportion of patients experience a relapse, and relapse rates are significantly higher in patients with pathogenic variants (Abu-Duhier et al., 2001; Dohner et al., 2010; Fathi and Chen, 2011; Frohling et al., 2002; Gilliland and Griffin, 2002; Levis et al., 2002;

Levis and Small, 2003; Nakao et al., 1996). Consequently, *FLT3* pathogenic variants are often associated with poor disease prognosis, and it is recommended that AML therapy be specified according to the patients' pathogenic variant status (O'Donnell et al., 2011).

Despite the role of *FLT3* in hematopoiesis and its association with poor prognosis, the complete clinical implications of germline pathogenic variants have not been clearly shown.

Here we demonstrate that the germline *FLT3* pathogenic variant is strongly associated with AML development and prognosis. We suggest that in addition to the somatic pathogenic variant, the germline pathogenic variant is a significant genetic factor affecting AML prognosis. In particular, the p.D358V pathogenic variant and the prognosis of AML patients were first reported in this study. This is a valuable finding to be used as a biomarker to predict the prognosis of AML patients throughout further studies.

MATERIALS AND METHODS

IRB approval

A total of 76 AML patients were enrolled in this study, and bone marrow and saliva samples were obtained from these patients. This study was approved by the Seoul National University Hospital Institutional Review Board (IRB approval number: 1201-099-396).

Sample preparation and DNA extraction

Human genomic DNA was extracted from saliva and bone marrow using the QuickGene DNA Whole Blood Kit S (KURABO INDUSTRY, Japan) according to the manufacturer's recommendations.

Exome sequencing

We sequenced exomes from genomic DNA using the Solexa sequencing technology platform (HiSeq 2000, Illumina, San Diego, CA, USA) following the manufacturer's instructions. Briefly, we randomly aliquoted 3 μg of genomic DNA using the Covaris System to generate about 150 bp inserts. The fragmented DNA was end-repaired using T4 DNA polymerase and Klenow polymerase, and Illumina paired-end adaptor oligonucleotides were ligated to the sticky ends. We analyzed the ligation mixture using agarose gel electrophoresis and purified fragments between 200 and 250 bp in size. To capture 50-Mb targeted exons, the purified DNA library was hybridized with the SureSelect Human All Exon V3 and V5 probe set (Agilent, Santa Clara, CA, USA) following the manufacturer's instructions. This captured exome library was then sequenced using the HiSeq 2000 platform according to the manufacturer's instructions.

Exome sequencing data analysis and validation through genotyping

We aligned the FASTQ sequence files to the human reference (human_g1k_v 37.fasta) using the Burrows-Wheeler Aligner (Li and Durbin, 2009) (BWA-0.7.5), producing a single SAM file. We converted the SAM file to BAM format using SortSam in Picard-tools-1.68. We then used chromosome sorting and marked PCR duplicates using Picard to eliminate the duplicates from subsequent analysis with the Genome Analysis Toolkit (GATK-1.6.5) (McKenna et al., 2010b). To identify germline and somatic pathogenic variants in the *FLT3* gene (chr13:28578189-28674647), we utilized the Haplotypecaller tool on GATK and the Broad Institute's MuTect, respectively. Functional annotation was performed using ANNOVAR (Wang et al., 2010), and we predicted the functional effect of germline pathogenic variants using Polyphen-2 (Adzhubei et al., 2013).

To validate our results, we genotyped the SNP rs34172843 (Thermo Fisher Scientific, South San Francisco, CA) in a total of 64 AML patients using a TaqMan assay (Livak, 1999). Briefly, approximately 20 ng of purified genomic DNA was added to a genotyping mixture consisting of 2X genotyping master mix, 20X SNP genotyping assay, DNase-free water, and template DNA, followed by a 40-cycle PCR reaction. Amplicons were analyzed using the ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA). Genotyping quality control was performed in 10% of the samples by assessing the concordance rate between duplicates (> 99.5%).

Characterization of D358V pathogenic variants in The Cancer Genome Atlas (TCGA) data set

Approval for access to the International Cancer Genome Consortium's (ICGC) Cancer Genome Atlas and clinical data was obtained as a Pan-Cancer Analysis of Whole Genomes (PCAWG) researcher (DACO-1063499). To analyze the distribution of

clinically significant D358V germline pathogenic variants in TCGA, SNP calling variant call format (VCF) files were downloaded from the National Cancer Institute (NCI) sftp server (sftp://dccsftp.nci.nih.gov). A total of 2,834 SNP VCF from patients with 41 cancer types were selected for further analysis. The somatic pathogenic variant profile of D358V pathogenic variants in TCGA was obtained from the cBioPortal database (http://cbioportal.org)

Comparison with normal population

To analyze the significance of *FLT3* germline pathogenic variants in AML, we compared the 76 AML patients recruited for this study with 2,504 individuals from the 1,000 Genomes Project (http://www.internationalgenome.org). Genotype information was extracted using tabix-0.2.6 (Li, 2011).

Statistical analysis

AML patients were classified according to the FAB classification system. Risk groups based on cytogenetic and molecular analyses of bone marrow samples were determined using the Medical Research Council (MRC) criteria (Grimwade et al., 2010). Patient demographic information and survival outcomes were obtained from medical records. Overall survival (OS) was defined as the duration from diagnosis to death from any cause, and relapse-free survival (RFS) was defined as the time from diagnosis until relapse or death from any cause. Living patients were censored at the date of the last contact. Kaplan-Meier plots were constructed to determine the clinical significance of D358V germline pathogenic variants, and Fisher's exact tests were performed using R studio.

RESULTS

Distribution of germline and somatic *FLT3* pathogenic variants in AML patients

To assess the distribution of germline and somatic *FLT3* pathogenic variants, we evaluated matched saliva and bone marrow samples from the 76 patients in our cohort using exome sequencing. Germline pathogenic variants were assessed from saliva samples, and somatic pathogenic variants were assessed in bone marrow tumor samples. The average patient age was 46 years, with 61% male and 39% female (Table 1-1).

We identified a total of four germline mutation loci on chromosome 13: 28624294, 28674628, 28622544, and 28608473. All of the mutations were not found in the tyrosine kinase domain; see Fig 1A. The most abundant germline pathogenic variant was p.T227M (chr13:28624294) found in 55 (72.3%) patients. The next most frequent pathogenic variant was p.D7G (chr13:28674628) found in 16 (19.7%) patients, followed by p.D358V (chr13:28622544) in 6 (7.8%) patients and p.V557I (chr13:28608473) in 1.3% of patients (Table 1-2).

Table legends

Table 1-1. Patient demographics

Table 1-2. Overall information of *FLT3* germline pathogenic variant in Korean AML cohort

Table 1-1

	Total	**D358V (+)**	**D358V (-)**
	(N=76)	**(N=6)**	**(N=70)**
Gender, N (%)			
Male	47 (61)	6 (100)	39 (55)
Female	29 (39)	0 (0)	31 (45)
Age (average)	46	58	45
FAB classification, N (%)			
AML-ET	1 (1)	0 (0)	1 (1)
AML-MDS	9 (12)	3 (50)	6 (8)
AML-MPD	1 (1)	0 (0)	1 (1)
M0	2 (3)	0 (0)	2 (3)
M1	7 (9)	0 (0)	7 (9)
M2	22 (29)	1 (17)	21 (30)
M3	13 (17)	2 (33)	11 (16)
M4	12 (16)	0 (0)	12 (16)
M5	3 (4)	0 (0)	3 (4)
Unknown	6 (8)	0 (0)	6 (8)

Risk Group, N (%)			
Adverse	23 (30.2)	3 (50)	20 (28.5)
Intermediate	21 (27.6)	1 (16.6)	20 (28.5)
Favorable	31 (40.7)	2 (33.3)	29 (41.4)
Unknown	1 (1.3)	0	1 (1.4)

Table 1-2

Gene	AA change	Pathogenic variant type	Chr	Position	Ref	Alt	rsID	Patients Frequency in AML cohort		Normal Asian Frequency in 1000G		Fisher Test		
								FLT3 mut	FLT3 wild	FLT3 mut	FLT3 wild	OR	P value	Polyphen-2
FLT3	T227M	MISSENSE	13	28624294	G	A	rs1933437	55	21	734	60	0.21	<0.001	0.004
FLT3	D7G	MISSENSE	13	28674628	T	C	No rsID	15	61	410	384	0.23	<0.001	0.451
FLT3	D358V	MISSENSE	13	28622544	T	A	rs34172843	6	70	25	769	2.63	4.00E-02	0.999
FLT3	V557I	MISSENSE	13	28608473	C	T	rs35958982	1	75	4	790	2.62	0.36	0

AA, amino acid; Chr, chromosome; Ref, reference; Alt, alternative; Mut, pathogenic variant; OR, odds ratio; CI, confidence interval

Two specific regions on the *FLT3* locus (p.D835V/H and p.D358V) were affected by somatic pathogenic variants (Fig 2). One pathogenic variant, in the p.D835 region, affected the *FLT3* tyrosine kinase domain, a frequently mutated locus in AML (Sheikhha et al., 2003). Two patients carried a p.D835 somatic pathogenic variant (p.D835V and p.D835H). A third pathogenic variant, p.D358V, affecting the *FLT3* immunoglobulin domain, was detected in a single patient. Unlike p.D835 pathogenic variants, p.D358V pathogenic variants are not well described in AML.

According to TCGA, the p.D358V somatic pathogenic variant is not reported for several other cancer types in addition to AML (Fig 1B). Interestingly, as described above, this locus was observed not only as a somatic pathogenic variant but also as a germline pathogenic variant in our AML cohort. Using polyphen-2 (Adzhubei et al., 2013) to predict the functional effect of the pathogenic variant, we determined that D358V is a deleterious pathogenic variant (polyphen score 0.999; see Table 1-2).

Figure legends

Figure 1-1. Distribution of *FLT3* somatic and germline pathogenic variants

Each bar represents a genetic variation. A) is germline pathogenic variant information which is observed in Korean AML patients. B) is somatic pathogenic variant patterns of the *FLT3* gene in The Cancer Genome Atlas (TCGA) data. Each bar represents a somatic pathogenic variant and the length of the bar means the number of pathogenic variants. The color of the bar indicates the type of pathogenic variant: the green color is the missense pathogenic variant; the black is the truncating pathogenic variant; the brown is the in-frame pathogenic variant; and the purple is the splicing region of the pathogenic variant. The D358V pathogenic variant was not reported in the TCGA database.

Figure 1-2. *FLT3* pathogenic variant characteristics

A plot is showing the pathogenic variant pattern of the *FLT3* gene screened in the Korean AML cohort. The blue marker is a patient with a somatic pathogenic variant, and the green marker is a patient with a germline pathogenic variant. Pathogenic variants at the 358 amino acid position of the *FLT3* gene (D358V) were found both as somatic pathogenic variants and as germ cell pathogenic variants in the Korean AML cohort.

Figure 1-1

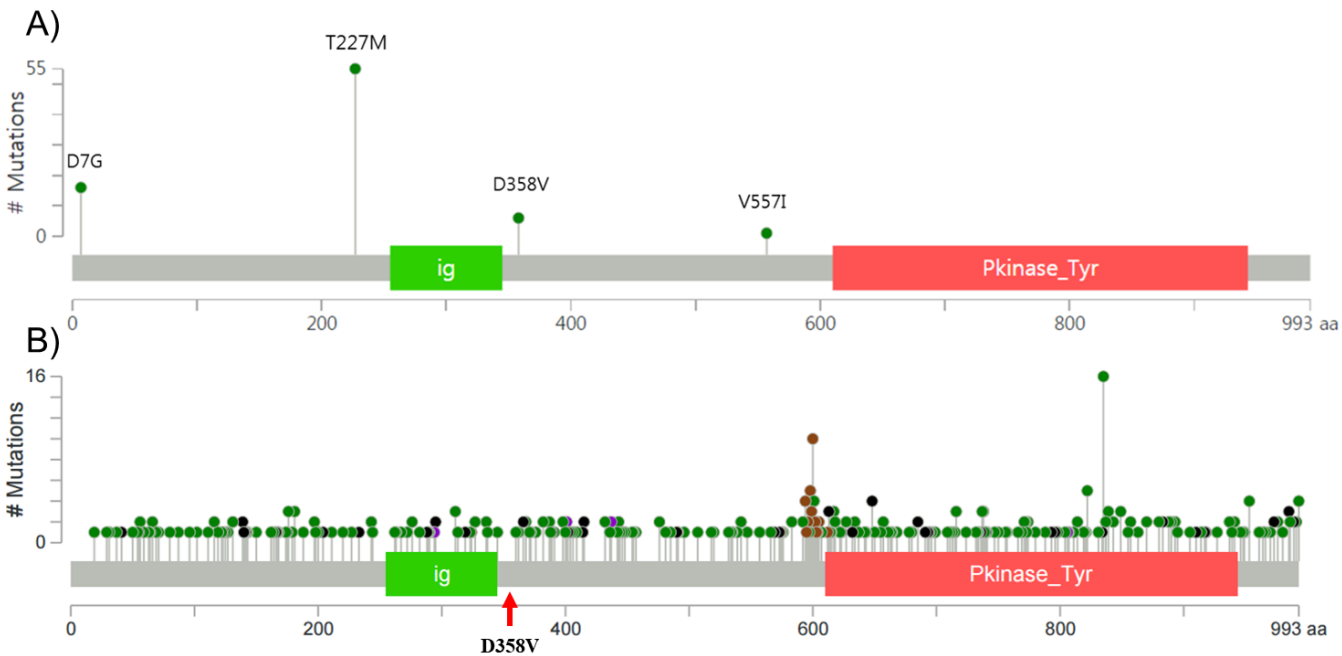
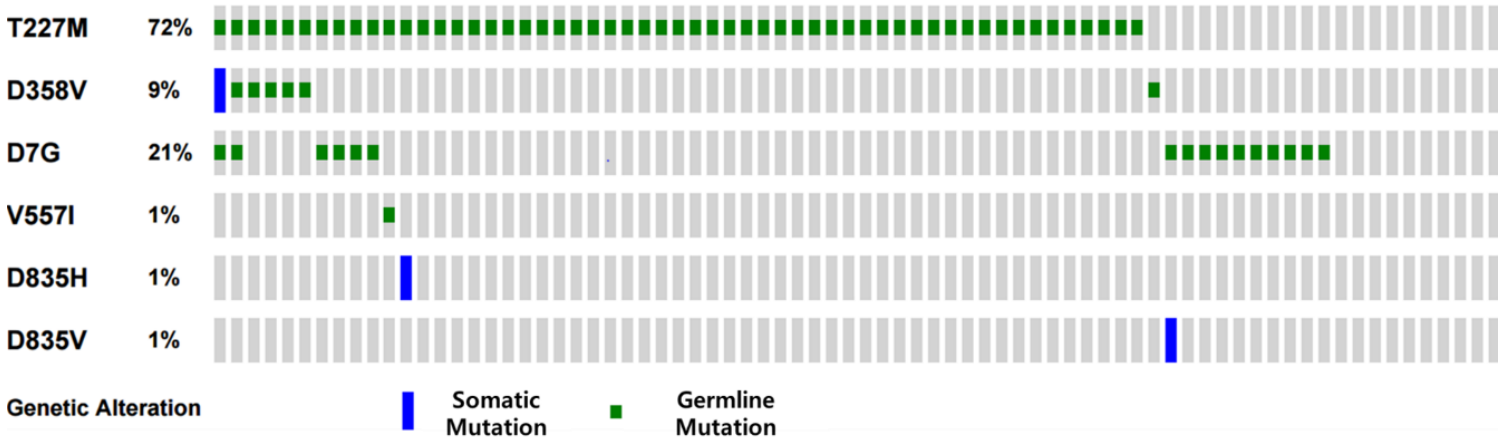

A)
T227M
55
Mutations
D7G
D358V
V557I
0
ig
Pkinase_Tyr
0
200
400
600
800
993 aa
B)
16
Mutations
0
ig
Pkinase_Tyr
0
200
400
600
800
993 aa
D358V

Figure 1-2


T227M
72%
D358V
9%
D7G
21%
V557I
1%
D835H
1%
D835V
1%
Genetic Alteration
Somatic Mutation
Germline Mutation

Prevalence and ethnic specificity of *FLT3* germline p.D358V pathogenic variants in AML

To determine whether *FLT3* germline pathogenic variants are associated with a predisposition to developing AML, we compared the frequency of *FLT3* germline pathogenic variants in our cohort to that of *FLT3* germline pathogenic variants among 754 members of the normal Asian population segment of the 1000 Genomes Project (n=794). The frequency of the p.D358V pathogenic variant was significantly higher (p=4.0E-02 and OR=2.63) in AML patients than in the general population, suggesting that a predisposition to AML development is associated, at least in part, with the *FLT3* germline pathogenic variant. In contrast, the T227M and D7G germline pathogenic variants were significantly higher in the normal Asian population (OR=0.21, p<0.001 and OR=0.23, p<0.001) than in AML patients. No significant difference in frequency was observed for the V557I (p=0.36) or pathogenic variants (Table 1-2) between AML patients and members of the general population.

Distribution of the p.D358V pathogenic variant in The Cancer Genome Atlas

To determine the distribution of the p.D358V *FLT3* pathogenic variant in TCGA, we collected a total of 834 SNP VCF called by the DFKZ

calling pipeline in 2,834 patients with 41 different types of cancer. We identified the p.D358V pathogenic variant in ten patients with five cancers: a US population with ovarian serous cystadenocarcinoma, a German population with pediatric brain cancer, a Japanese population with liver cancer, a Korean population with acute myeloid leukemia, and a Chinese population with gastric cancer. The p.D358V had the highest observed frequency in the Japanese liver cancer cohort (2.4%; 6 of 250 patients).

Clinical implications of the p.D358V germline pathogenic variant

We classified AML patients, according to the FAB classification and determined risk groups using the MRC criteria. The most common FAB classification subtype was M2 (22%), and most patients were classified into the favorable risk group (40.7%). There were no significant differences in clinical characteristics between the two groups.

Overall survival was significantly reduced in patients with the p.D358V pathogenic variant (p=0.03 by log-rank test; see Fig 3A). Disease-free survival also decreased, though the difference was not significant (p=0.52 by log-rank test).

To further assess the prognostic value of p.D358V in AML, we investigated the correlation between p.D358V and survival outcomes in the Japanese liver cancer cohort from TCGA. The overall survival of

liver cancer patients was significantly reduced in patients with p.D358V (p=0.03; see Fig 3B), suggesting that the p.D358V germline pathogenic variant affects prognosis across several cancer types. The association between *FLT3* and liver cancer is not well known, but records from the cBioPortal database indicated twelve of 1,115 patients (1.1%) with hepatocellular carcinoma had *FLT3* mutations. Interestingly, seven of them were truncation mutations that lost *FLT3* gene function, suggesting that the loss-of-function of *FLT3* would benefit liver cancer cell growth. The overall survival of patients with hepatocellular carcinoma was significantly reduced (p=0.01; see Figure 1-4) in patients with the *FLT3* mutation, similar to the outcome of patients with D358V.

In summary, the decreased activity of the *FLT3* gene may be beneficial for the survival of the tumor cell and lead to a poor prognosis for the patient.

Figure legends

Figure 1-3. Clinical significance of the germline *FLT3* D358V pathogenic variant

Prognosis of patients according to D358V genetic variation: (+) is the overall survival of patients with genetic pathogenic variants, and (-) is the overall survival of patients without genetic pathogenic variants. A) is the overall survival for the liver cancer cohort of the AML cohort and B) is that for the ICGC data set.

Figure 1-4. Clinical significance of the *FLT3* somatic mutation in liver cancer

Cases with alterations are shown with a red line. On the other hand, cases without alterations are shown with a blue line.

Figure 1-5. White blood cell count according to the *FLT3* D358V pathogenic variant

Each dot represents a white blood cell count from an individual patient. White blood cell counts are compared between patients with and without the p.D358V mutation.

Figure 1-3

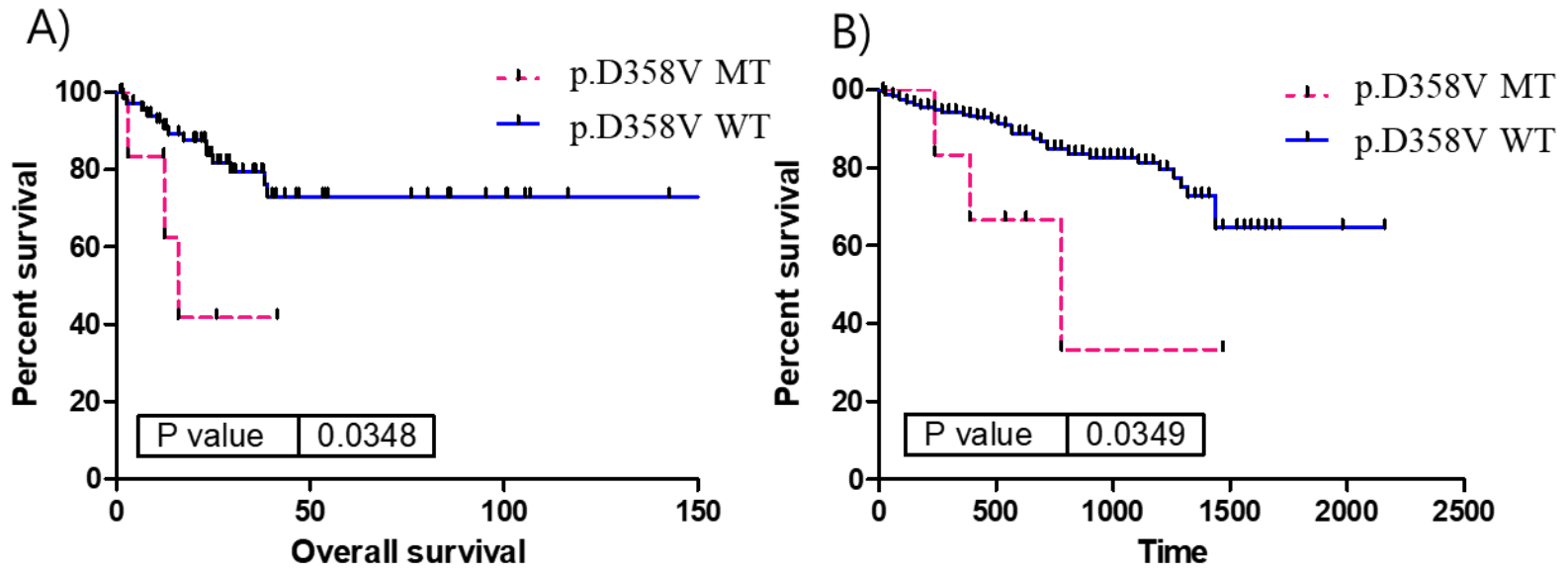

A)
p.D358V MT
p.D358V WT
Percent survival
100
80
60
40
20
0
P value
0.0348
0
50
100
150
Overall survival
B)
p.D358V MT
p.D358V WT
Percent survival
00
80
60
40
20
0
P value
0.0349
0
500
1000
1500
2000
2500
Time

Figure 1-4

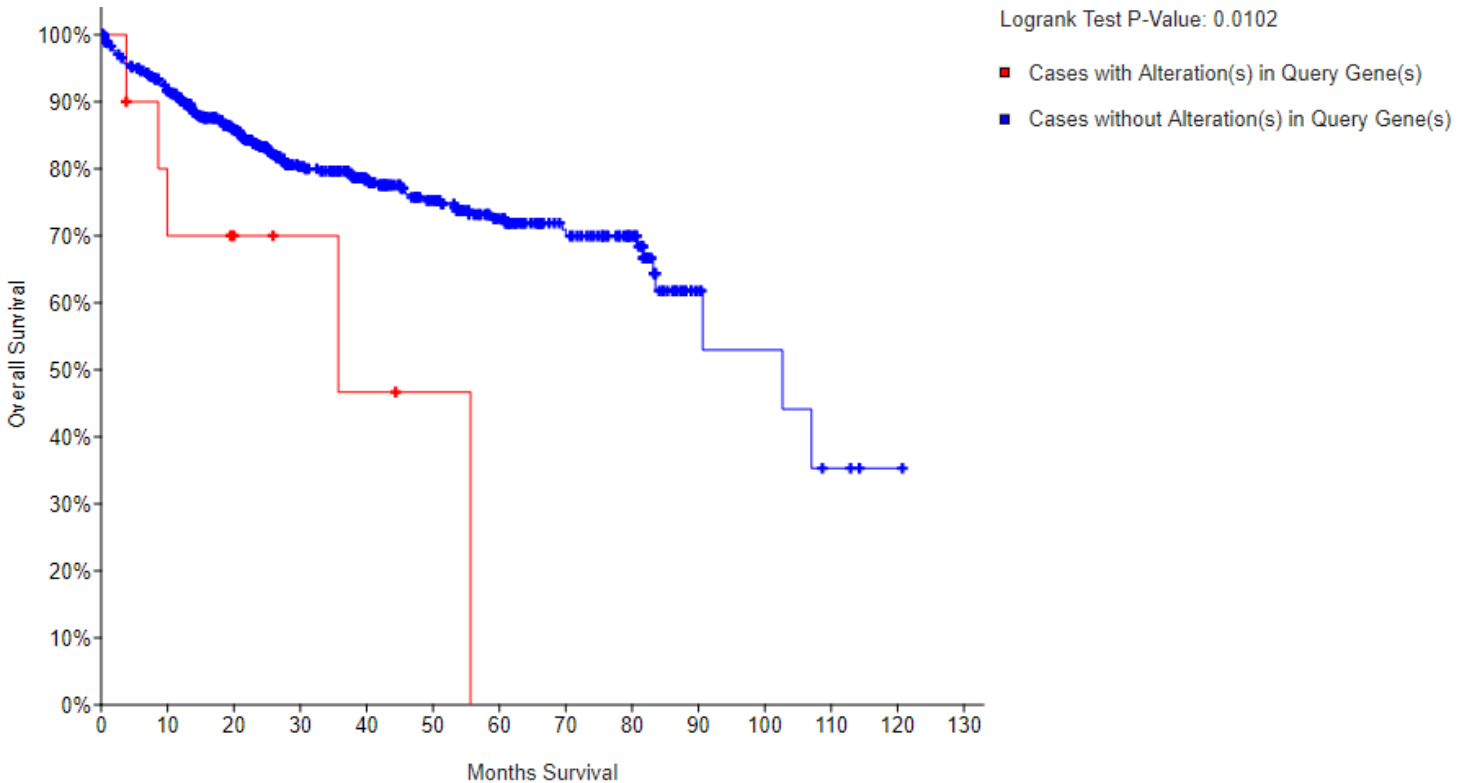
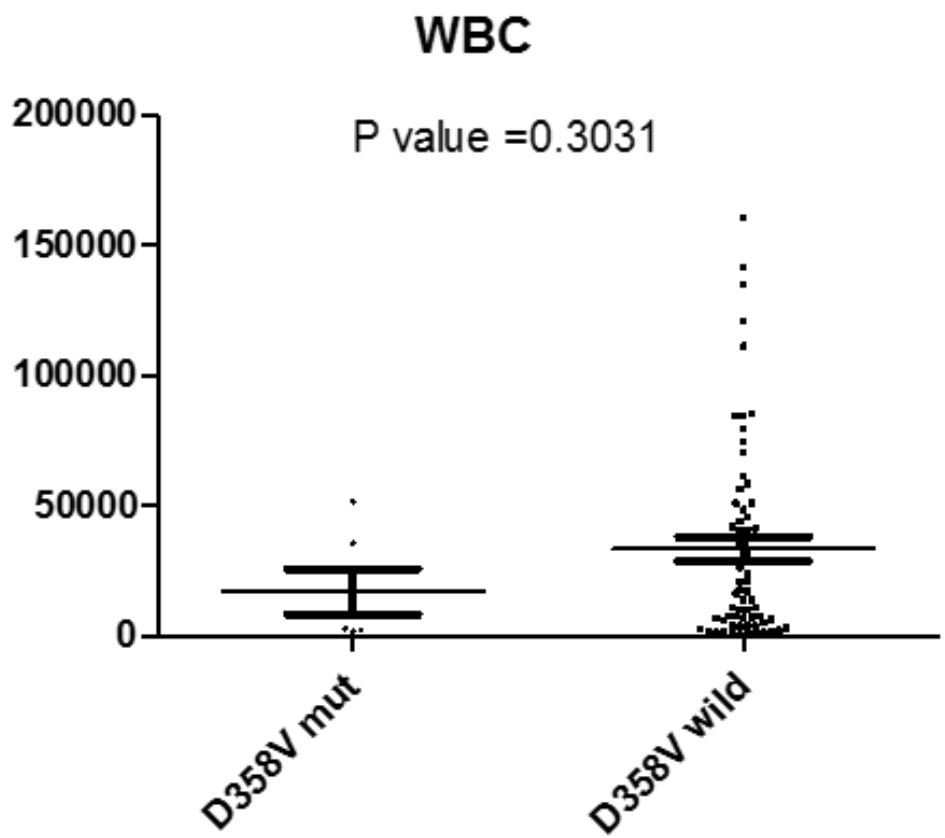

Logrank Test P-Value: 0.0102
Cases with Alteration(s) in Query Gene(s)
Cases without Alteration(s) in Query Gene(s)
Overall Survival
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
10
20
30
40
50
60
70
80
90
100
110
120
130
Months Survival

Figure 1-5


WBC
P value =0.3031
200000
150000
100000
50000
0
D358V mut
D358V wild

In general, patients with TKD pathogenic variants or ITD pathogenic variants, which are a functional gain of the *FLT3* gene, have higher white blood cell counts compared to normal individuals. To estimate how the p.D358V pathogenic variant plays a role in the activity of the *FLT3* gene, we compared the white blood cell counts between patients with the pathogenic variant and without the pathogenic variant. Comparing white blood cell counts at the time of diagnosis with gene activity (How et al., 2012) revealed no difference in white blood cell counts between patients with or without the p.D358V germline pathogenic variant; however, the white blood cell count had a tendency to decrease in patients with p.D358V, implying that this pathogenic variant induces low gene activity (Figure 1-5).

DISCUSSION

AML-specific *FLT3* pathogenic variants have been an intense research focus over the past few decades. The effects of somatic pathogenic variants, which usually present as the highly prevalent activating pathogenic variants such as *FLT3* TKD or *FLT3* ITD, are well understood. In contrast, the effect of germline *FLT3* pathogenic variants remains largely unknown, as does their impact on prognosis. Here, while screening a Korean AML cohort, we found an enrichment in a three germline *FLT3* pathogenic variants (p.D358V, p.T227M, and p.D7G). Although the presence of two alleles (T227M and D7G) was related to a reduction in the risk for AML ($P<0.001$, respectively), the clinical significance was not confirmed in this cohort. Nielka P et al. reported that the presence of the p.T227M allele reduced the risk of leukopenia, but the clinical effect of the mutation was not assessed. In addition, p.D7G was identified in CML patients with BCR-ABL fusion but not biochemically characterized; thus, the effects on FLT3 protein function were not yet known (Lyu et al., 2016; van Erp et al., 2009). In contrast, the p.D358V pathogenic variant was associated with poor disease prognosis.

Germline pathogenic variants of tyrosine kinase receptors such as *KIT* or *EGFR* have been reported to contribute to tumorigenesis in other

cancers (Centeno et al., 2011; Hirota et al., 2002). An *FLT3* germline variant (R387Q) was identified to have an overlapping somatic pathogenic variant in endometrial cancer (Lu et al., 2015). Most pathogenic variants affecting tyrosine kinase receptors such as *KIT*, *EGFR*, and *FLT3* are known to induce tumors by increasing gene activity through self-activation or by disabling gene regulation [24-26]. In contrast, Sandhofer et al. suggested that the loss-of-function phenotype of the p.Q569Vfs pathogenic variant may induce cell proliferation in AML by activating alternative cell-signaling independent of activity [27]. It was out of the scope of our work to elucidate the biological significance of the *FLT3* p.D358V germline pathogenic variant we identified in AML patients. The pathogenic variant does not alter the white blood cell count significantly; however, it tends to reduce white blood cell counts. In addition, we confirmed through *in silico* analysis that the pathogenic variant blocks *FLT3* enzyme activity, which supports the finding that this variant affects low gene activity. While we cannot exclude the possibility that the p.D358V pathogenic variant is a natural single nucleotide polymorphism with no influence on leukemogenesis, it is important to note that germline pathogenic variants in the same locus have been detected repetitively in tumor cells. Thus, we can infer that this pathogenic variant facilitates the growth of leukemia cells by escaping *FLT3*-mediated regulation of cell growth.

To elucidate the mechanism of receptor interaction with the p.D358V pathogenic variant, functional characterization of the *FLT3* pathogenic variant is critical. We found that *FLT3* pathogenic variants did not always activate the gene; therefore, we recommend that the functional characteristics of *FLT3* pathogenic variants should be investigated individually. While such analyses were outside of the scope of this study, we provide an important description of the *FLT3* germline pathogenic variant and report for the first time the association between the p.D358V pathogenic variant and disease prognosis.

CHAPTER 2

Association between mild primary immunodeficiency disease and cancer risk

INTRODUCTION

Immune deficiency diseases, which are associated with immune system failure, are classified into two main groups. Primary immunodeficiency diseases (PID) are caused by congenital genetic disorders, whereas secondary immunodeficiency diseases (SID) result from acquired factors, such as human immunodeficiency virus (HIV) infection or organ transplants.

Due to the abnormal function of the immune system, patients with both types of immunodeficiency exhibit various clinical symptoms, including persistent or recurring infections, auto-inflammatory disorders, tumors, and disorders of various organs. Among these symptoms, tumor cell development results from the suppression of immune surveillance. The theory of immune surveillance states that the immune system is constantly monitoring the presence of cancerous cells, and in most cases, it will destroy these abnormal cells. However, in patients with decreased immune surveillance activity due to PID or SID, abnormal cells can escape and proliferate to produce tumors. The association between cancer development and decreased immune surveillance function has been well-established, and several immunodeficiency diseases are known to increase cancer incidence. Thus, there is strong epidemiological evidence for the role of immune surveillance in cancer

prevention, as evidenced by the increased risk of cancers detected among specific immunocompromised patients (Mayor et al., 2018). In particular, various cancers often occur in patients with acquired immunodeficiency syndrome (AIDS) due to HIV infection. Some of the most common carcinomas associated with AIDS include non-Hodgkin lymphoma, Kaposi's sarcoma, and cervical cancer (Yarchoan and Uldrick, 2018).

An increased risk of secondary solid cancers among allogeneic hematopoietic cell transplant (HCT) recipients is also well-documented. HCT recipients have significantly higher than expected rates of invasive solid cancers, such as oral cavity, esophageal, lung, soft tissue, and brain cancer (Majhail et al., 2011). In addition, solid organ transplant recipients have elevated risks for cancer development due to both immunosuppression and oncogenic viral infections. Thus, recipients of a kidney, liver, heart, or lung transplant experience higher rates of a diverse array of infection-related and unrelated cancers, as compared with the general population (Engels et al., 2011).

Unlike SIDs, which arise due to external influences, such as infection or transplantation, PIDs are caused by a defect in various genes. As with SIDs, an association between PIDs and various cancers has been noted for decades (Filipovich et al., 1992; Hayward et al., 1997; Mayor et al., 2018). The increased cancer incidence observed in PID patients has been studied using a multi-center registry of subjects that focuses on

those with clinically overt immunodeficiency. This cohort mostly includes patients with diseases that are epidemiologically classified as autosomal recessive, which implies a loss-of-function due to the double allele defect.

However, some of the patients in this study were diagnosed with an autosomal dominant form of PID. The main cause of this phenomenon is known as haploinsufficiency, which was first reported in 1989 for a condition called angioedema (Ariga et al., 1989). Haploinsufficiency refers to the occurrence of a particular mutant phenotype in heterozygotes, despite a lack of negative dominance of the mutant allele over its wild-type (WT) counterpart. Genes with haploinsufficiency show incomplete penetrance when considering the predicted 50% residual function of the gene product (Rieux-Laucat and Casanova, 2014). Thus, phenotypes can range from mild to moderate, depending on the extent of gene penetration, and this can result in the presence of heterozygous carriers with mild traits in the normal population. Critically, in these individuals, the altered function of the mutated gene may impair cellular homeostasis and promote abnormal cell growth, which can increase cancer risk.

This phenomenon was noted in a previous study that revealed a correlation between heterozygous carrier status for lysosomal storage disease (LSD) and cancer (Shin et al., 2018). Notably, the risk of cancer

was found to be significantly increased in heterozygous carriers of 42 LSD-related gene variants. Importantly, this study was possible through two international collaborative studies, the International Cancer Genome Consortium (ICGC) and the 1000 Genomes Project, both of which produced large-scale whole genome sequencing (WGS) data. These two datasets contain genetic information from unrelated adult individuals without severe pediatric disease, whose genomes were sequenced as part of various cancer-specific and population genetic studies. Therefore, pathogenic variants identified from these studies are considered to have a low penetration of Mendelian disease, and correlations can be estimated by comparing the variant distribution between the two datasets.

In this study, we hypothesize that in mild carriers of PID-associated alleles, the accumulation of abnormal cells, which are not adequately removed by defective immune surveillance mechanisms in these individuals, results in increased risk for the development of malignant cancer. Here we report a comprehensive analysis of the pan-cancer association for putative pathogenic germline mutations in 114 PID-related genes using data from the ICGC and the 1000 Genomes Project. Our data indicate an increased risk for the development of certain cancers in carriers of pathogenic gene variants. Moreover, the reliability of this phenomenon was confirmed by validating this result in an external cohort.

MATERIALS AND METHODS

Study cohort

We used two large whole genome databases, each containing a subset of donors who were not diagnosed with immunodeficiency disease until they reached adulthood. The ICGC dataset contains genetic information from 2,566 individuals with cancer (Pan-Cancer cohort), and the control dataset from the 1000 Genomes Project (1000 Genomes cohort) is comprised of 2,504 genomes (The Genomes Project, 2015).

A total of 1,667 patients, who were diagnosed with bladder cancer (n=530), liver cancer (n=465), hematology cancer (n=297), or stomach cancer (n=375), was used as an external validation cohort for target sequencing. Recruitment for the external validation cohort of this study was approved by the Seoul National University Hospital Institutional Review Board (IRB approval number: 1805-067-946).

Putative pathogenic variant selection (PPV)

After an extensive literature review, we identified 114 PID-related genes (Table 2-1). First of all, we identified a truncated mutation which was expected to cause functional deleterious substitution (Stop-gain/loss, Start-loss, Splice site) and frameshift indel. We then screened curated disease databases, including the Human Gene Mutation Database

Professional 2018.2 (HGMD) and ClinVar, for pathogenic variants with possible deleterious function. In this way, we identified putative disease-causing variants that present strong evidence for functional deleteriousness. Based on the evidence that most allele frequencies (AF) for pathogenic variants in the clinical database are less than 0.1%, we excluded all variants with mean AFs (arithmetic means of the Pan-Cancer and 1000 Genomes cohorts) >0.1% from the putative pathogenic variant (PPV) selection process (Figure 2-1). An overall flow chart for our PPV selection method is shown in Figure 2-2.

Figure Legends

Figure 2-1. The Exome Aggregation Consortium (ExAC) allele frequencies for pathogenic variants of PID-related genes identified in the ClinVar and HGMD databases

The ExAC allele frequencies of variants described as pathogenic mutations in 114 PID-related genes are divided into five ranges, and the number of variants in each range is plotted.

Figure 2-2. Overall flow chart for our PPV selection method

Table 2-1. The 114 genes associated with primary immunodeficiency disease identified in this study

Figure 2-1

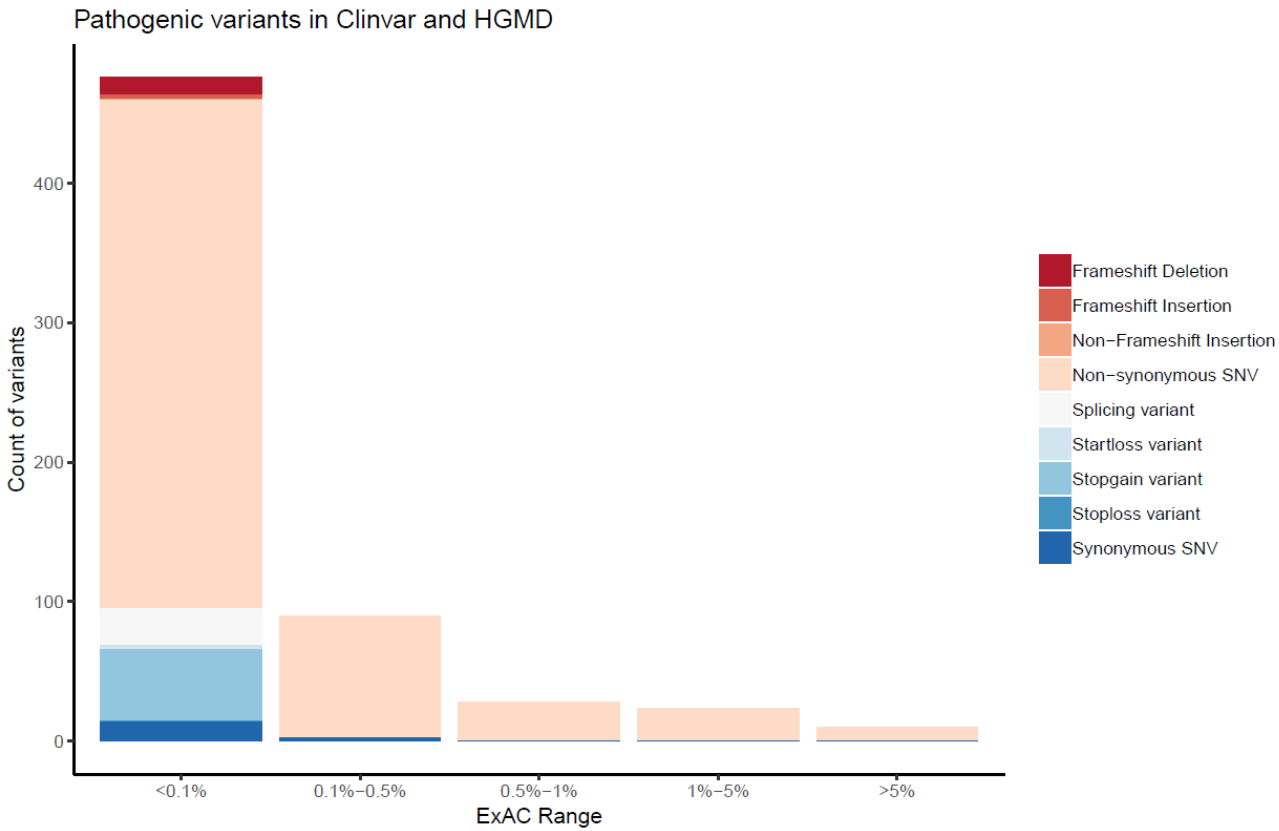

Pathogenic variants in Clinvar and HGMD
Count of variants
400
300
200
100
0
<0.1%
0.1%–0.5%
0.5%–1%
1%–5%
>5%
ExAC Range
Frameshift Deletion
Frameshift Insertion
Non–Frameshift Insertion
Non–synonymous SNV
Splicing variant
Startloss variant
Stopgain variant
Stoploss variant
Synonymous SNV

Figure 2-2

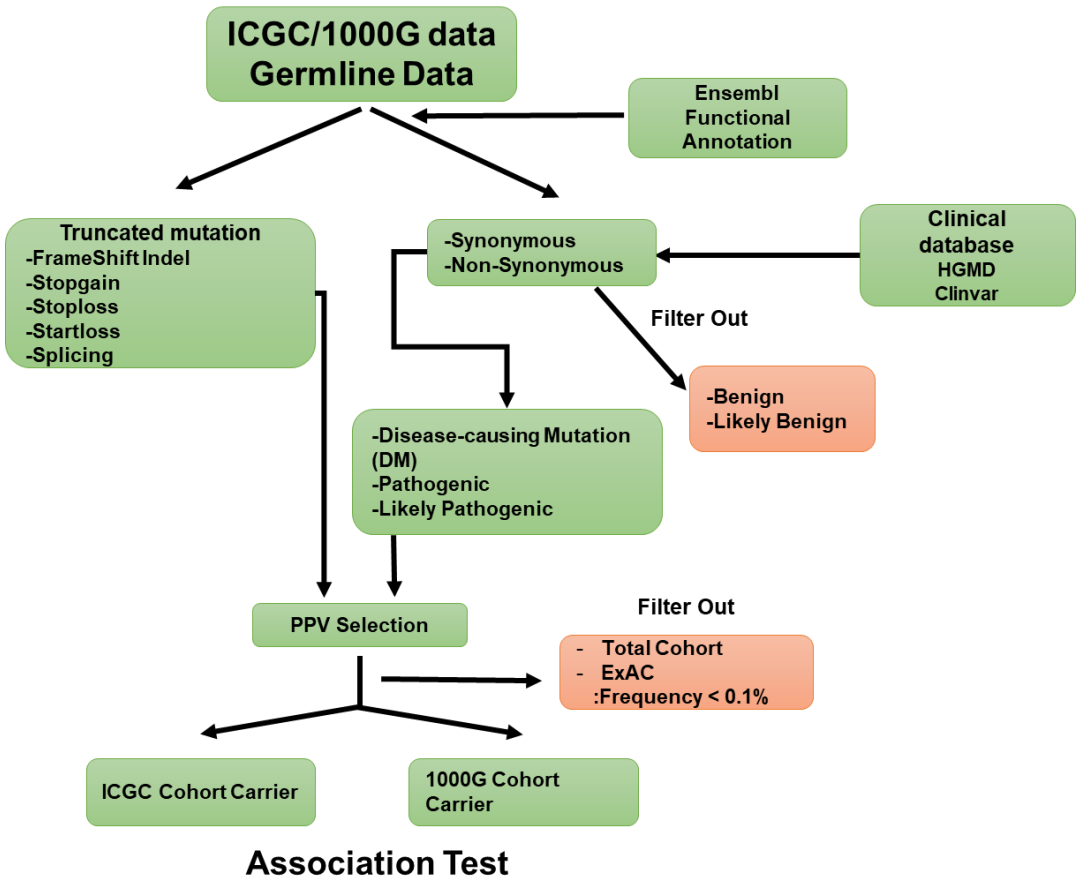

ICGC/1000G data
Germline Data
Ensembl
Functional
Annotation
Truncated mutation
-FrameShift Indel
-Stopgain
-Stoploss
-Startloss
-Splicing
-Synonymous
-Non-Synonymous
Clinical
database
HGMD
Clinvar
Filter Out
-Benign
-Likely Benign
-Disease-causing Mutation
(DM)
-Pathogenic
-Likely Pathogenic
PPV Selection
Filter Out
- Total Cohort
- ExAC
:Frequency < 0.1%
ICGC Cohort Carrier
1000G Cohort
Carrier
Association Test

Table 2-1

Gene	Phenotype	Inheritance	Classification
DOCK8	Hyper-IgE recurrent infection syndrome	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
IRF7	Immunodeficiency 39	Autosomal Recessive	Defects in intrinsic and innate immunity
ELANE	Neutropenia	Autosomal Dominant	Congenital defects of phagocyte
ISG15	Immunodeficiency 38	Autosomal Recessive	Defects in intrinsic and innate immunity
TNFRSF4	Immunodeficiency 16	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
TERT	Pulmonary fibrosis and/or bone marrow failure	Autosomal Dominant	CID with associated or syndromic features
CD81	Immunodeficiency 6	Autosomal Recessive	Predominantly antibody deficiencies
CARD11	B-cell expansion with NFκB and T-cell anergy, immunodeficiency 11A, immunodeficiency 11B with atopic dermatitis	Autosomal Dominant	Predominantly antibody deficiencies
TRNT1	retinitis pigmentosa, and erythrocytic microcytosis, Sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay	Autosomal Recessive	Predominantly antibody deficiencies
MEFV	Familial Mediterranean fever	Autosomal Dominant	Undefined
STIM1	Immunodeficiency 10	Autosomal Recessive	CID with associated or syndromic features
CD27	Lymphoproliferative syndrome 2	Autosomal Recessive	Diseases of immune dysregulation
AICDA	Immunodeficiency with hyper-IgM	Autosomal Recessive	Predominantly antibody deficiencies
PIK3CD	Immunodeficiency 14	Autosomal Dominant	Predominantly antibody deficiencies
IL17RC	Candidiasis	Autosomal Recessive	Defects in intrinsic and innate immunity
TYK2	Immunodeficiency 35	Autosomal Recessive	Defects in intrinsic and innate immunity
PARN	Dyskeratosis congenita, pulmonary fibrosis and/or bone marrow failure	Autosomal Dominant	CID with associated or syndromic features
DCLRE1C	Omenn syndrome, severe combined immunodeficiency	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity

TNFRSF13B	Immunodeficiency 2	Autosomal Recessive	Predominantly antibody deficiencies
IL17RA	Immunodeficiency 51	Autosomal Recessive	Defects in intrinsic and innate immunity
JAK3	Severe combined immunodeficiency	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
IL12RB1	Immunodeficiency 30	Autosomal Recessive	Defects in intrinsic and innate immunity
RFXANK	MHC class II deficiency	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
PNP	Immunodeficiency due to purine nucleoside phosphorylase deficiency	Autosomal Recessive	CID with associated or syndromic features
IGLL1	Agammaglobulinemia 2	Autosomal Recessive	Predominantly antibody deficiencies
TINF2	Dyskeratosis congenita, Revesz syndrome	Autosomal Dominant	CID with associated or syndromic features
FOXN1	T-cell immunodeficiency, congenital alopecia, and nail dystrophy	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
UNC119	Immunodeficiency 13	Autosomal Dominant	Immunodeficiencies affecting cellular and humoral immunity
IL21R	Immunodeficiency 56	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
EXTL3	Immunoskeletal dysplasia with neurodevelopmental abnormalities	Autosomal Recessive	CID with associated or syndromic features
CD19	Immunodeficiency, common variable, 3	Autosomal Recessive	Predominantly antibody deficiencies
LAT	Immunodeficiency 52	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CORO1A	Immunodeficiency 8	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
TCN2	Transcobalamin II deficiency	Autosomal Recessive	CID with associated or syndromic features
DNMT3B	Immunodeficiency-centromeric instability-facial anomalies syndrome 1	Autosomal Recessive	CID with associated or syndromic features
LCK	Immunodeficiency 22	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
TAP2	Bare lymphocyte syndrome, type I	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
TAP1	Bare lymphocyte syndrome, type I	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity

TAPBP	Bare lymphocyte syndrome, type I	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
IFNAR2	Immunodeficiency 45	Autosomal Recessive	Defects in intrinsic and innate immunity
NOP10	Dyskeratosis congenita, autosomal recessive 1	Autosomal Recessive	CID with associated or syndromic features
IFNGR2	Immunodeficiency 28, mycobacteriosis	Autosomal Recessive	Defects in intrinsic and innate immunity
IL7R	Severe combined immunodeficiency, T-cell negative, B-cell/natural killer-cell positive type	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
NFKBIA	Ectodermal dysplasia and immunodeficiency	Autosomal Dominant	CID with associated or syndromic features
RAG2	Combined cellular and humoral immune defects with granulomas, Omenn syndrome, severe combined immunodeficiency, B-cell negative	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CSF3R	Neutropenia, severe congenital, 7, autosomal recessive	Autosomal Recessive	Congenital defects of phagocyte
NCF4	Granulomatous disease, chronic, autosomal recessive, cytochrome b-positive, type III	Autosomal Recessive	Congenital defects of phagocyte
RAC2	Neutrophil immunodeficiency syndrome	Not Determined	Congenital defects of phagocyte
MYD88	Pyogenic bacterial infections	Not Determined	Defects in intrinsic and innate immunity
RHOH	Epidermodysplasia verruciformis	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
STAT5B	Growth hormone insensitivity with immunodeficiency	Not Determined	CID with associated or syndromic features
CTPS1	Immunodeficiency 24	Autosomal Recessive	Diseases of immune dysregulation
IKBKB	Immunodeficiency 15A, immunodeficiency 15B	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
G6PC3	Dursun syndrome, neutropenia, severe congenital 4	Autosomal Recessive	Congenital defects of phagocyte
TNFRSF13C	Immunodeficiency, common variable, 4	Autosomal Recessive	Predominantly antibody deficiencies
ADA	Severe combined immunodeficiency	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
EPG5	Vici syndrome	Autosomal Recessive	CID with associated or syndromic features
IRAK4	Invasive pneumococcal disease, recurrent isolated, 1	Not Determined	Defects in intrinsic and innate immunity
RELB	Immunodeficiency 53	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity

BLOC1S6	Hermansky-Pudlak syndrome 9	Autosomal Recessive	CID with associated or syndromic features
ITGB2	Leukocyte adhesion deficiency	Autosomal Recessive	Congenital defects of phagocyte
TTC7A	Gastrointestinal defects and immunodeficiency syndrome	Autosomal Recessive	CID with associated or syndromic features
PRKDC	Immunodeficiency 26	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
IKZF1	Immunodeficiency, common variable, 13	Autosomal Dominant	Predominantly antibody deficiencies
IL17F	Candidiasis	Autosomal Dominant	Defects in intrinsic and innate immunity
RAB27A	Griscelli syndrome, type 2	Autosomal Recessive	Diseases of immune dysregulation
MALT1	Immunodeficiency 12	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
STAT2	Immunodeficiency 44	Autosomal Recessive	Defects in intrinsic and innate immunity
MS4A1	Immunodeficiency, common variable, 5	Autosomal Recessive	Predominantly antibody deficiencies
CD79B	Agammaglobulinemia 6	Autosomal Recessive	Predominantly antibody deficiencies
RTEL1	Dyskeratosis congenita	Autosomal Dominant	CID with associated or syndromic features
FERMT3	Leukocyte adhesion deficiency, type III	Autosomal Recessive	Congenital defects of phagocyte
PIK3R1	Agammaglobulinemia 7, immunodeficiency 36, SHORT syndrome	Autosomal Recessive	Predominantly antibody deficiencies
AP3B1	Hermansky-Pudlak syndrome 2	Autosomal Recessive	Diseases of immune dysregulation
PGM3	Immunodeficiency 23	Autosomal Recessive	CID with associated or syndromic features
BCL10	Immunodeficiency 37	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
IRF8	Immunodeficiency 32A, mycobacteriosis, immunodeficiency 32B, monocyte and dendritic cell deficiency	Autosomal Recessive	Defects in intrinsic and innate immunity
CD8A	CD8 deficiency	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CYBA	Chronic granulomatous disease	Autosomal Recessive	Congenital defects of phagocyte
NBN	Nijmegen breakage syndrome	Autosomal Recessive	CID with associated or syndromic features
GFI1	Neutropenia, severe congenital 2	Autosomal Dominant	Congenital defects of phagocyte

HELLS	Immunodeficiency-centromeric instability-facial anomalies syndrome 4	Autosomal Recessive	CID with associated or syndromic features
BLNK	Agammaglobulinemia 4	Autosomal Recessive	Predominantly antibody deficiencies
BCL11B	Immunodeficiency 49	Autosomal Dominant	Immunodeficiencies affecting cellular and humoral immunity
NFKB1	Immunodeficiency, common variable, 12	Autosomal Dominant	Predominantly antibody deficiencies
NFKB2	Immunodeficiency, common variable, 10	Autosomal Dominant	Predominantly antibody deficiencies
ATM	Ataxia-telangiectasia	Autosomal Recessive	CID with associated or syndromic features
UNG	Immunodeficiency with hyper-IgM, type 5	Autosomal Recessive	Predominantly antibody deficiencies
CD3E	Immunodeficiency 18	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CD3D	Immunodeficiency 19	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CD3G	Immunodeficiency 17, CD3 gamma-deficient	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
ORAI1	Immunodeficiency 9, myopathy, tubular aggregate, 2	Autosomal Dominant	CID with associated or syndromic features
IL21	Immunodeficiency, common variable, 11	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
GATA2	Emberger syndrome, immunodeficiency 21	Autosomal Dominant	Congenital defects of phagocyte
POLE	FILS syndrome	Autosomal Recessive	CID with associated or syndromic features
IFNGR1	Immunodeficiency 27A/B, mycobacteriosis	Autosomal Recessive	Defects in intrinsic and innate immunity
CARD9	Candidiasis	Autosomal Recessive	Congenital defects of phagocyte
VPS45	Neutropenia, severe congenital 5	Autosomal Recessive	Congenital defects of phagocyte
RFX5	Bare lymphocyte syndrome, type II, complementation group C/E	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
HAX1	Neutropenia, severe congenital 3	Autosomal Recessive	Congenital defects of phagocyte
ITK	Lymphoproliferative syndrome 1	Autosomal Recessive	Diseases of immune dysregulation
FCGR3A	Immunodeficiency 20	Autosomal Recessive	Defects in intrinsic and innate immunity

CD247	Immunodeficiency 25	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
DOCK2	Immunodeficiency 40	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CDCA7	Immunodeficiency-centromeric instability-facial anomalies syndrome 3	Autosomal Recessive	CID with associated or syndromic features
WIPF1	Wiskott-Aldrich syndrome 2	Not Determined	CID with associated or syndromic features
NHP2	Dyskeratosis congenita, autosomal recessive 2	Autosomal Recessive	CID with associated or syndromic features
NCF2	Chronic granulomatous disease due to deficiency of NCF-2	Autosomal Recessive	Congenital defects of phagocyte
STAT1	Immunodeficiency 31A, immunodeficiency 31B, immunodeficiency 31C	Autosomal Dominant	Defects in intrinsic and innate immunity
PTPRC	Severe combined immunodeficiency, T-cell negative, B-cell/natural killer-cell positive	Autosomal Recessive	Immunodeficiencies affecting cellular and humoral immunity
CR2	Immunodeficiency, common variable, 7	Autosomal Recessive	Predominantly antibody deficiencies
NHEJ1	Severe combined immunodeficiency with microcephaly, growth retardation, and sensitivity to ionizing radiation	Not Determined	Immunodeficiencies affecting cellular and humoral immunity
SP110	Hepatic venoocclusive disease with immunodeficiency	Autosomal Recessive	CID with associated or syndromic features
LYST	Chediak-Higashi syndrome	Autosomal Recessive	Diseases of immune dysregulation

*CID: Combined immunodeficiency, MHC: Major Histocompatibility Complex,

Adjustment for population structure differences

Principal component analysis (PCA) was performed using the individual genotypes of tag single-nucleotide polymorphisms (tag-SNPs), which represent the population-distinguishing linkage disequilibrium blocks. We identified a total of 1,555,886 candidates for tag-SNPs from MAP and PED files that were downloaded from the phase 3 HapMap ftp server (ftp://ftp.ncbi.nlm.nih.gov/hapmap/phase_3/). We then converted the genomic positions of these tag-SNPs to GRCh37/hg19 coordinates using the Batch Coordinate Conversion (liftOver) tool, created by the UCSC Genome Browser Group (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

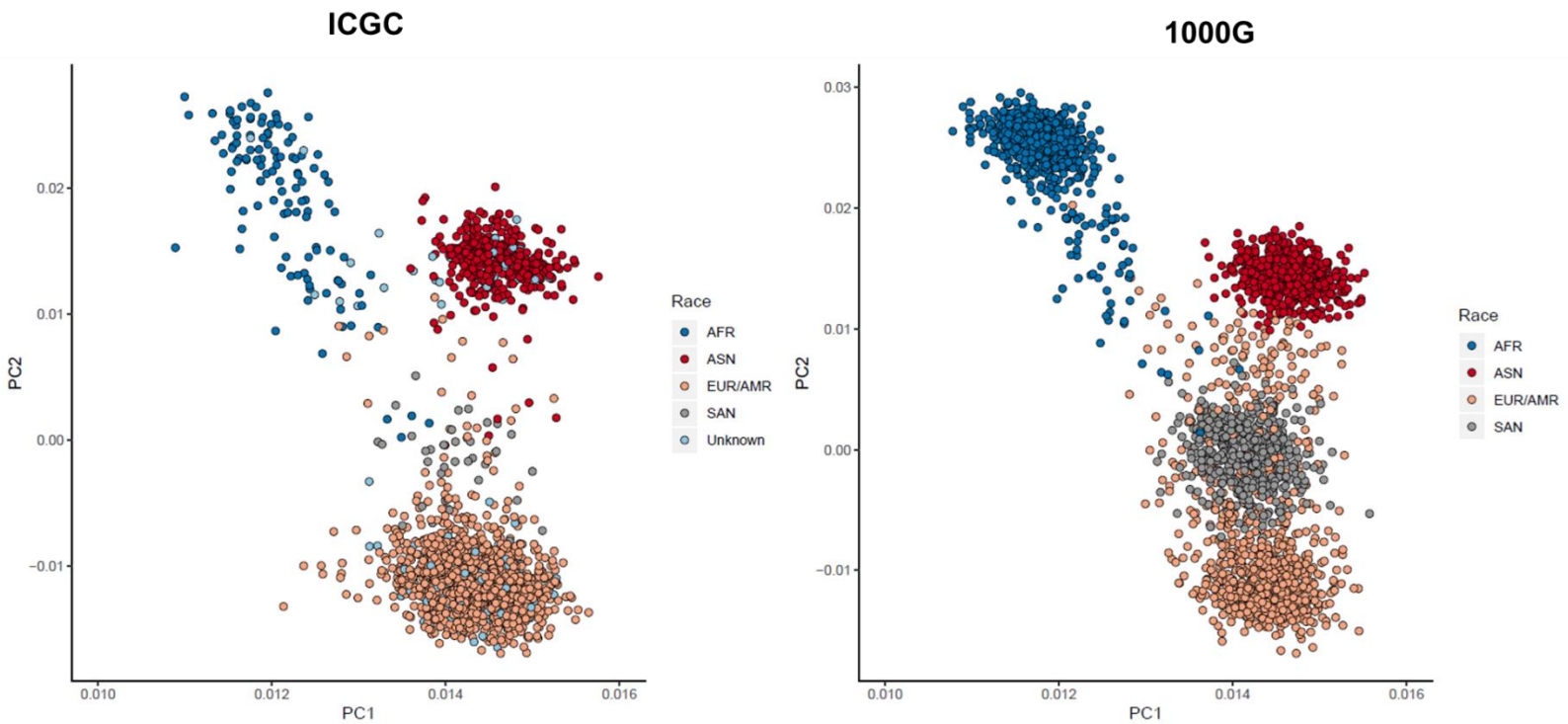
The Genome Analysis Toolkit was used to merge VCF files from the Pan-Cancer and 1000 Genomes cohorts (McKenna et al., 2010a), and tag-SNPs with AFs ≥5% and ≤50% were extracted from merged VCF files using VCFtools, version 1.13, resulting in 16,304 filtered tag-SNPs (Danecek et al., 2011). Among these, population-stratifying tag-SNPs were selected using the PLINK pruning method (Purcell et al., 2007), thereby reducing linkage disequilibrium panels containing a number of SNPs to a condensed subset of representative tag-SNPs (Novembre et al., 2008). As a result, 10,494 tag-SNPs were selected for use in the subsequent PCA (Figure 2-3).

Figure Legend

Figure 2-3. PC1 and PC2 values generated from PCA analysis using tag-SNPs from the Pan-Cancer and 1000 Genomes cohorts

Subjects are classified into four groups based on ethnicity: African (AFR), East Asian (ASN), South Asian (SAN), and European/American (EUR/AMR), and the PC1 and PC2 values for each group are compared. These values reflect genetic differences in the four groups and validate the values for population structure adjustment.

Figure 2-3


ICGC
1000G
PC1
PC2
Race
AFR
ASN
EUR/AMR
SAN
Unknown
0.010
0.012
0.014
0.016
0.03
0.02
0.01
0.00
−0.01

Statistical analysis

The Pan-Cancer and 1000 Genomes cohorts were analyzed using the SKAT-O test for rare-variant association, the Fisher Exact test, and logistic regression for direct comparison of mutation prevalence (Lee et al., 2012). We also adjusted for population structure using PCA on the 10,494 filtered tag-SNPs (Figure 2-3). Wilcoxon rank-sum tests and logistic regression were performed for between-group comparisons of age, and CIBERSORT was used for immune cell milieu analysis (Chen et al., 2018).

Correction for multiple testing was conducted using the false discovery rate (FDR) estimation procedure, and the tail area-based FDR was reported (Benjamini et al., 2001). All tests were two-tailed unless otherwise specified. We considered an FDR <0.1 or *P*-value <0.05 (when not adjusted for multiple testing) to be statistically significant. Identification of differentially expressed genes (DEGs) and gene set analysis were performed using the DESeq2 (Love et al., 2014) Bioconductor package and bioinformatics enrichment tools (Huang da et al., 2009), based on the framework of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway maps (Kanehisa and Goto, 2000), R version 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria), in combination with pipelines from Bioconductor version 3.5, were used for all analyses.

Mutation signature analysis

Statistical comparison of mutational load (logistic regression) between PPV-bearing and PPV-free tumors for each mutational signature showed no significant results at a threshold of $P<0.05$. Data used in this analysis were from the official beta version of mutational signatures for single nucleotide substitutions generated and uploaded to the SFTP server (sftp://dccsftp.nci.nih.gov/pancan/) by the Pan-Cancer Analysis of Whole Genomes Project team in March 2017.

Target sequencing for validation

DNA was extracted from 400 μl of blood or buffy coat using the Exgene Blood SV kit (GeneAll Biotechnology, Seoul, Korea). DNA samples (500 ng) were then sheared into 250–300 bp fragments with a Q800R3 Sonicator (Qsonica, Newtown, CT, USA). The ACCEL-NGS 2S DNA Library Kit was used to create index-ligated libraries according to the LIBRARY KITS (Swift Biosciences, Ann Arbor, MI, USA) protocol, and the dsDNA library concentrations were measured with the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). Ten libraries containing 150-ng DNA per sample were pooled and subjected to hybridization according to the xGen hybridization capture DNA libraries (IDT) protocol. Briefly, xGen Lockdown Probes (43 genes or 144 genes) were incubated at 65°C for 4 or 16 hours, and streptavidin

beads were used for capture. Samples were sequenced using an Illumina HiSeq 2500 sequencer, using a read length of 2×100 bp.

RESULTS

Description of study cohorts

The Pan-Cancer cohort is comprised of four ethnic populations and 38 histological subtypes of pediatric or adult cancer. We graphed the population distribution of this cohort, as well as the control 1000 Genomes cohort for comparison, and found that EUR/AMR was the most common ethnicity in the Pan-Cancer ICGC cohort (Figure 2-4). The 1000 Genomes normal control cohort is comprised of four ethnic classifications, with EUR/AMR being the most prevalent (The Genomes Project, 2015).

PPV prevalence in the Pan-Cancer and 1000 Genomes cohorts

For the 114 PID-associated genes, we identified 36,292 germline single nucleotide variants (SNVs) and small insertions and deletions (indels) in the variant call sets of the Pan-Cancer and 1000 Genomes cohorts. Of these, 502 PPVs were defined through the selection process described above. The overall distribution of PPV carriers among cancer patients was 13.6%, whereas the distribution of PPV carriers in normal controls was 8.5%. This is statistically significant after adjusting for population structure (Odds ratio [OR]: 1.8, confidence interval [CI]: 1.5–2.2) (Figure 2-5). Conversely, we detected no difference in the distribution of

synonymous variants that do not result in functional changes, thus confirming that the PPV distribution in the two groups was not due to chance or structural bias.

Across the 30 major histological subtypes (each with >15 individuals per subtype), PPV prevalence ranged from 5.5% to 33.3%, and for seven of these, the PPV prevalence was significantly higher than that in the 1000 Genomes cohort. These were bladder cancer (21.7%), breast adenocarcinoma (13.3%), liver cancer (13.0%), chronic lymphocytic leukemia (CLL; 17.7%), pancreas adenocarcinoma (17.6%), pancreas endocrine (24.6%), and soft tissue leiomyoma (33.3%).

We then classified the 114 PID genes into seven categories based on the immunological symptoms of their associated disease (Bousfiha et al., 2018; Machet, 1998) (Figure 2-6) and determined the PPV prevalence for each group in the two cohorts. Among these seven classifications, PPV enrichment in five (i.e., Diseases of immune dysregulation, Congenital defects of phagocytes, Immunodeficiencies affecting cellular and humoral immunity, Predominantly antibody deficiencies, and Combined immunodeficiency disease (CID) with associated or syndromic features) was significantly associated with increased cancer incidence (Figure 2-7-a). The highest frequency classification was “CID with associated or syndromic features.” We performed clustering analysis using the frequency of PPV in each

histology type. Clustering analysis of the PPV distribution pattern for each disease classification further revealed dominant patterns of PPV enrichment in “CID with associated or syndromic features” and “Predominantly antibody deficiencies” (Figure 2-7-b).

Figure Legends

Figure 2-4. Demographics of the study cohorts

The pie chart on the left is a representation of the percentage of ethnographic classifications for each of the two cohorts. The 2,566 subjects from the ICGC (Pan-Cancer) cohort are shown on the upper left, and the 2,504 individuals from the 1000 Genomes cohort are on the lower left. The right bar chart indicates the ethnicity information for each histological subtype in the ICGC cohort.

Figure 2-5. Comparison of PPV prevalence in each cohort

The bar chart on the left shows PPV frequency for the 114 PID-associated genes in the Pan-Cancer (ICGC) and control 1000 Genomes (TG) cohorts. The right bar chart shows the odds ratios (ORs) calculated from the comparison between each histological subtype and TG. *P*-values were calculated by logistic regression and are expressed in different colors: grey >0.05, yellow <0.05, orange <0.01, and red <0.001.

Figure 2-6. Functional classification of the diseases associated with the 114 PID genes

The chart on the left shows the distribution ratios for the 114 PID genes within the seven immune classifications. Additional information on the

subclasses (not including “undefined”) is listed on the right.

Figure 2-7. Comparison of PPV prevalence within each functional classification

a) A comparison of PPV prevalence within the seven immune classifications of the Pan-Cancer (ICGC) and 1000 Genomes (1000G) cohorts is shown. The top graph shows the odds ratios (ORs) for each classification, with the frequencies graphed below. b) A clustering analysis based on the PPV frequency within the seven classifications is shown. Scaled values are expressed for each histological subtype. Red is assigned a high frequency PPV and blue is a low frequency PPV.

Figure 2-4

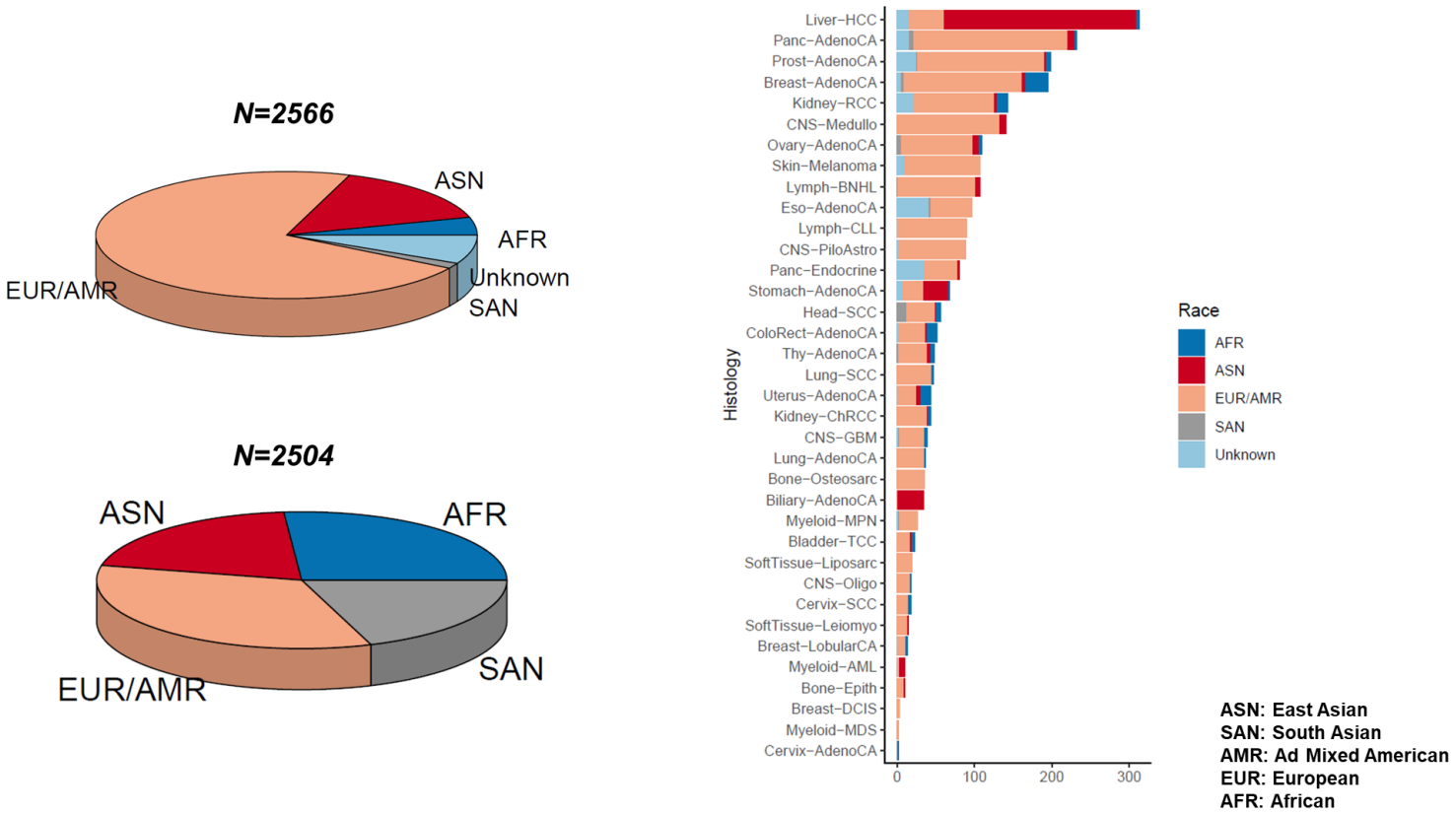

N=2566
ASN
AFR
Unknown
SAN
EUR/AMR
N=2504
ASN
AFR
SAN
EUR/AMR
Liver−HCC
Panc−AdenoCA
Prost−AdenoCA
Breast−AdenoCA
Kidney−RCC
CNS−Medullo
Ovary−AdenoCA
Skin−Melanoma
Lymph−BNHL
Eso−AdenoCA
Lymph−CLL
CNS−PiloAstro
Panc−Endocrine
Stomach−AdenoCA
Head−SCC
ColoRect−AdenoCA
Thy−AdenoCA
Lung−SCC
Uterus−AdenoCA
Kidney−ChRCC
CNS−GBM
Lung−AdenoCA
Bone−Osteosarc
Biliary−AdenoCA
Myeloid−MPN
Bladder−TCC
SoftTissue−Liposarc
CNS−Oligo
Cervix−SCC
SoftTissue−Leiomyo
Breast−LobularCA
Myeloid−AML
Bone−Epith
Breast−DCIS
Myeloid−MDS
Cervix−AdenoCA
Histology
0
100
200
300
Race
AFR
ASN
EUR/AMR
SAN
Unknown
ASN: East Asian
SAN: South Asian
AMR: Ad Mixed American
EUR: European
AFR: African

Figure 2-5

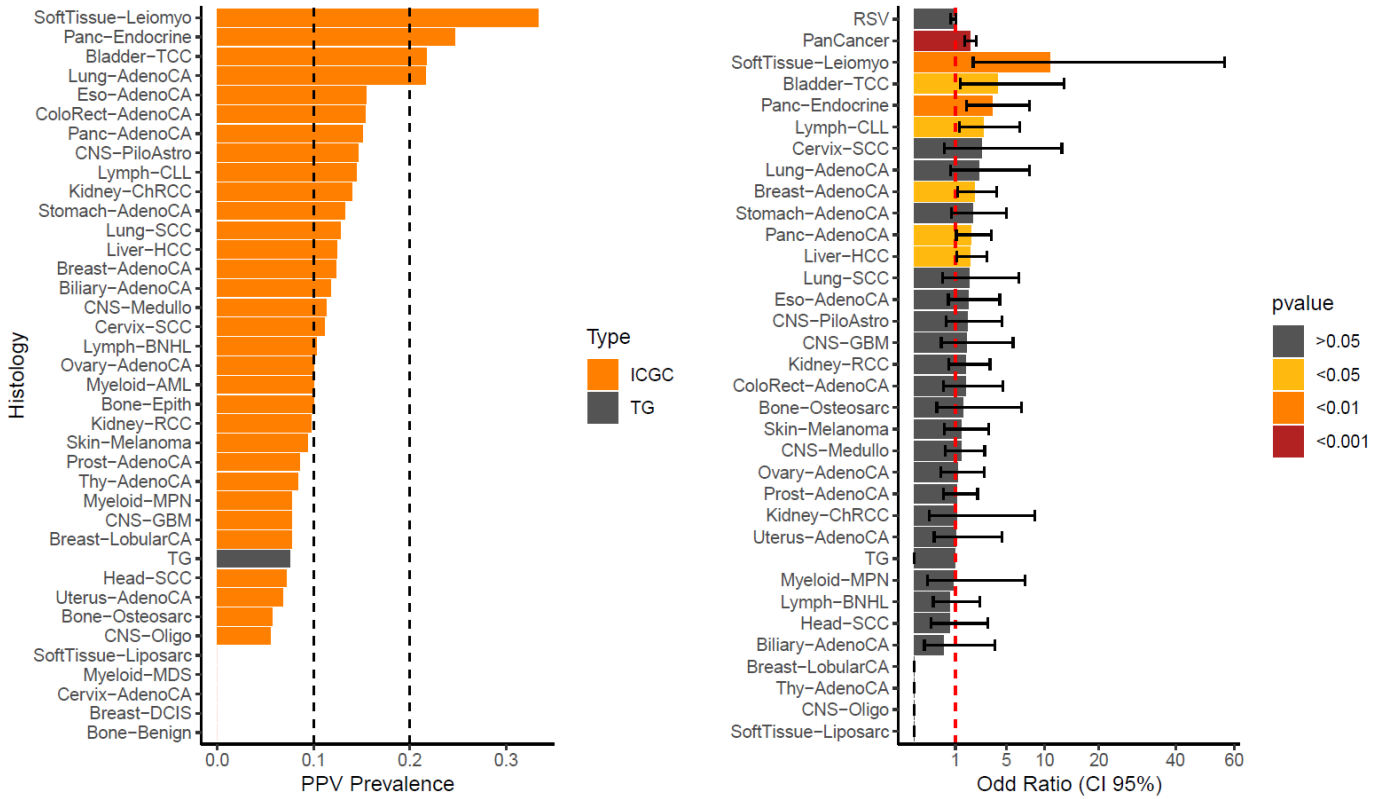
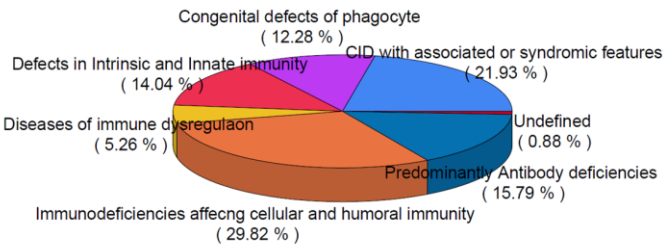

SoftTissue−Leiomyo
Panc−Endocrine
Bladder−TCC
Lung−AdenoCA
Eso−AdenoCA
ColoRect−AdenoCA
Panc−AdenoCA
CNS−PiloAstro
Lymph−CLL
Kidney−ChRCC
Stomach−AdenoCA
Lung−SCC
Liver−HCC
Breast−AdenoCA
Biliary−AdenoCA
CNS−Medullo
Cervix−SCC
Lymph−BNHL
Ovary−AdenoCA
Myeloid−AML
Bone−Epith
Kidney−RCC
Skin−Melanoma
Prost−AdenoCA
Thy−AdenoCA
Myeloid−MPN
CNS−GBM
Breast−LobularCA
TG
Head−SCC
Uterus−AdenoCA
Bone−Osteosarc
CNS−Oligo
SoftTissue−Liposarc
Myeloid−MDS
Cervix−AdenoCA
Breast−DCIS
Bone−Benign
Histology
0.0
0.1
0.2
0.3
PPV Prevalence
Type
ICGC
TG
RSV
PanCancer
SoftTissue−Leiomyo
Bladder−TCC
Panc−Endocrine
Lymph−CLL
Cervix−SCC
Lung−AdenoCA
Breast−AdenoCA
Stomach−AdenoCA
Panc−AdenoCA
Liver−HCC
Lung−SCC
Eso−AdenoCA
CNS−PiloAstro
CNS−GBM
Kidney−RCC
ColoRect−AdenoCA
Bone−Osteosarc
Skin−Melanoma
CNS−Medullo
Ovary−AdenoCA
Prost−AdenoCA
Kidney−ChRCC
Uterus−AdenoCA
TG
Myeloid−MPN
Lymph−BNHL
Head−SCC
Biliary−AdenoCA
Breast−LobularCA
Thy−AdenoCA
CNS−Oligo
SoftTissue−Liposarc
1
5
10
20
40
60
Odd Ratio (CI 95%)
pvalue
>0.05
<0.05
<0.01
<0.001

Figure 2-6



Defects in Intrinsic and Innate immunity

Bacterial and Parasitic Infections
MSMD and Viral infection

Immunodeficiencies affecting cellular and humoral immunity

Severe combined immunodeficiencies SCID, defined by CD3 T cell lymphopenia
Combined Immunodeficiencies Generally Less Profound than Severe Combined Immunodeficiency

Congenital defects of phagocyte

Neutropenia
Functional defects

CID with associated or syndromic features

Congenital thrombocytopenia
DNA Repair Defects

Predominantly Antibody deficiencies

Hypogammaglobulinemia
Other Antibody deficiencies

Diseases of immune dysregulaon

Hemophagocyc Lymphohisocytosis HLH & EBV suscepbility
Sd with Autoimmunity and Others

Figure 2-7-a

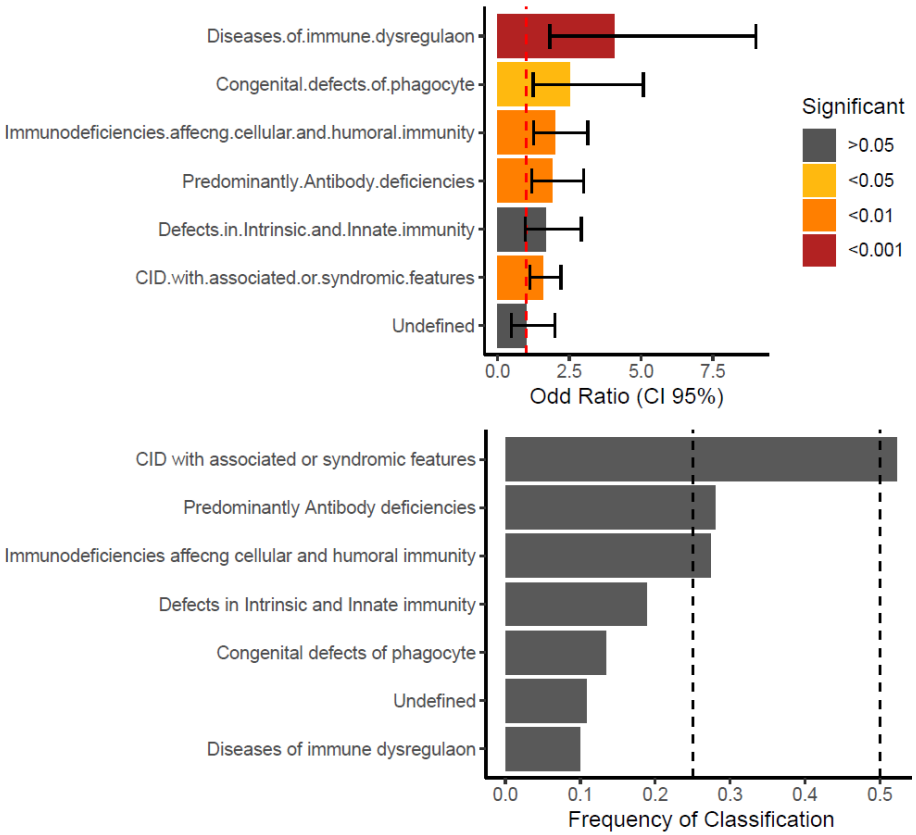
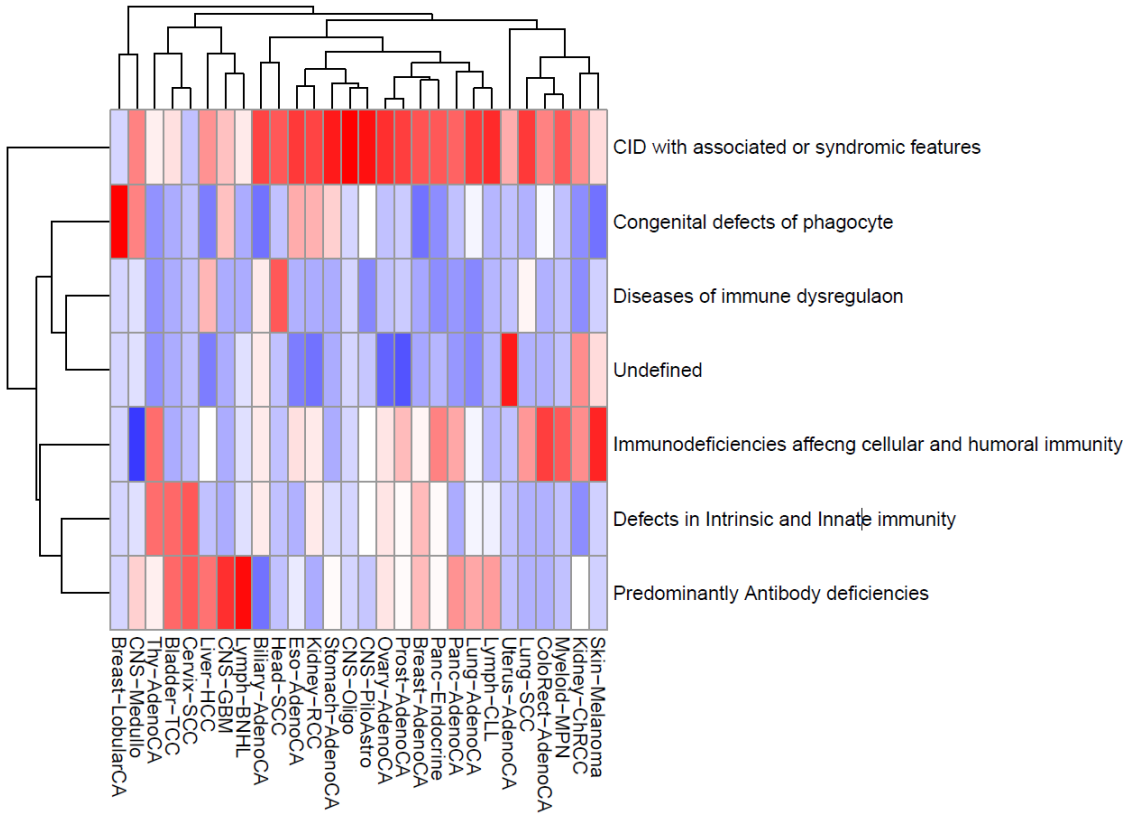

Diseases.of.immune.dysregulaon
Congenital.defects.of.phagocyte
Immunodeficiencies.affecng.cellular.and.humoral.immunity
Predominantly.Antibody.deficiencies
Defects.in.Intrinsic.and.Innate.immunity
CID.with.associated.or.syndromic.features
Undefined
0.0
2.5
5.0
7.5
Odd Ratio (CI 95%)
Significant
>0.05
<0.05
<0.01
<0.001
CID with associated or syndromic features
Predominantly Antibody deficiencies
Immunodeficiencies affecng cellular and humoral immunity
Defects in Intrinsic and Innate immunity
Congenital defects of phagocyte
Undefined
Diseases of immune dysregulaon
0.0
0.1
0.2
0.3
0.4
0.5
Frequency of Classification

Figure 2-7-b


CID with associated or syndromic features
Congenital defects of phagocyte
Diseases of immune dysregulaon
Undefined
Immunodeficiencies affecng cellular and humoral immunity
Defects in Intrinsic and Innate immunity
Predominantly Antibody deficiencies
Breast−LobularCA
CNS−Medullo
Thy−AdenoCA
Bladder−TCC
Cervix−SCC
Liver−HCC
CNS−GBM
Lymph−BNHL
Biliary−AdenoCA
Head−SCC
Eso−AdenoCA
Kidney−RCC
Stomach−AdenoCA
CNS−Oligo
CNS−PilоAstro
Ovary−AdenoCA
Prost−AdenoCA
Breast−AdenoCA
Panc−Endocrine
Panc−AdenoCA
Lung−AdenoCA
Lymph−CLL
Uterus−AdenoCA
Lung−SCC
ColoRect−AdenoCA
Myeloid−MPN
Kidney−ChRCC
Skin−Melanoma

Association between PPV-carrier status and histological specificity

We then measured the association between each histological subtype and PPV-carrier status for each of the 114 PID genes and found that 82 cancer-gene pairs were significantly associated by SKAT-O analysis, as adjusted for the population structure (FDR<0.1) (Figure 2-8 and Table 2-2). Overall, 27 histological subclasses were significantly associated with at least one gene, whereas 59 PID genes were associated with at least one cancer type. The PPV prevalence in gene-histology pairs ranged from 0.3% to 8.6%, and the pair with the highest prevalence was the *TNFRSF13B* gene in bladder cancer.

Age of cancer onset according to PPV-carrier status

To better understand the effects of PID on cancer development, we next compared the age of cancer onset according to PPV-carrier status. We found that 18 histological subtypes showed a tendency toward decreased age of onset associated with PPV-carrier status. Of these, PPV carriers showed a significantly reduced onset age for bladder cancer (median age, 49 vs. 69 years; P=0.002 by a one-sided test) and liver cancer (median age, 60 vs. 66 years; P=0.002) (Figure 2-9-a).

The liver cancer groups within the ICGC include the US project, LIHC-US, which is mostly composed of non-viral hepatocellular carcinoma, and the Japanese project, LIRI-JP, which is primarily

composed of viral hepatocellular carcinoma. Interestingly, a more significant association between PPV-carrier status and reduced onset age was obtained for the LIRI-JP cohort (*P*-value for viral-associated hepatocellular carcinoma = 0.02; *P*-value for non-viral-associated hepatocellular carcinoma = 0.18; Figure 2-9-b). Among viral-associated hepatocellular carcinoma, the hepatitis B virus (HBV) tended to promote an earlier onset than the hepatitis C virus (HCV).

The LIRI-JP cohort obtained from Japanese patients is composed primarily of virus-related liver cancer patients and has the largest composition with a homogeneous dataset. We therefore focused on this cohort for further analysis, as it is optimized to observe PID-related carcinogenesis mechanisms and to better detect the association between PID and virus-related cancer.

An analysis of single gene effects revealed that PPVs in *MEFV* and *TNFRSF13B* were significantly associated with reduced age of onset for liver and bladder cancer, respectively. When the PID genes were classified by associated immune pathology, PPV carriers in the "Predominantly antibody deficiencies" category showed a significantly decreased age of onset for both liver and bladder cancer, whereas those in "CID with associated or syndromic features" and "undefined" showed a decreased age of onset only for liver cancer (Figure 2-10-a,b).

Figure Legend

Figure 2-8. Association between each histological subtype and PPV-carrier status for each of the 144 PIDD genes

The association for each gene-histology pair is represented by the dot plot of the false discovery rate (FDR), calculated by the SKAT-O test. The size of the dot is the number of patients with the gene, and the color of the dot is assigned to the degree of –log (FDR).

Table 2-2. Statistically significant gene-histology pairs determined by the SKAT-O test

Figure 2-9. Association between the age of cancer onset and PPV-carrier status

A box plot comparing the age of cancer onset for PPV carriers vs. non-carriers is shown. The *P*-values at the top of the figure are calculated from the one-sided Wilcoxon test, and those shown in red are statistically significant: a) age of onset in PPV carriers vs. non-carriers for each histological subtype and b) age of onset in PPV carriers vs. non-carriers for individual liver cancer groups within the Pan-Cancer ICGC cohort.

Figure 2-10. Association between the age of onset for liver and

bladder cancer and PPV-carrier status within individual PID genes and immune pathology classifications

The ages of onset for liver and bladder cancer, which are significantly different, are compared in PPV carriers vs. non-carriers, both within specific genes and in the six groups of genes classified by immune pathology: a) age of onset analysis for liver cancer and b) age of onset analysis for bladder cancer.

Figure 2-8

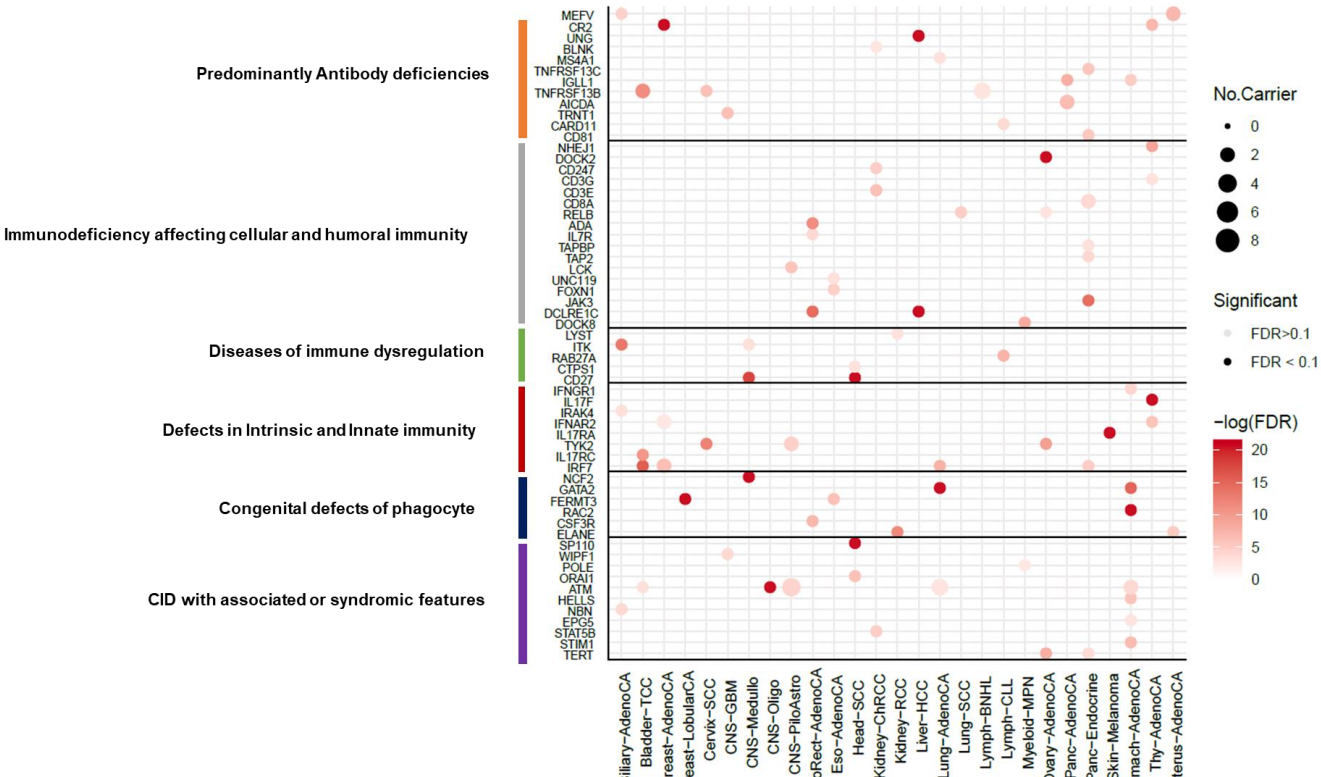

Predominantly Antibody deficiencies
Immunodeficiency affecting cellular and humoral immunity
Diseases of immune dysregulation
Defects in Intrinsic and Innate immunity
Congenital defects of phagocyte
CID with associated or syndromic features
MEFV
CR2
UNG
BLNK
MS4A1
TNFRSF13C
IGLL1
TNFRSF13B
AICDA
TRNT1
CARD11
CD81
NHEJ1
DOCK2
CD247
CD3G
CD3E
CD8A
RELB
ADA
IL7R
TAPBP
TAP2
LCK
UNC119
FOXN1
JAK3
DCLRE1C
DOCK8
LYST
ITK
RAB27A
CTPS1
CD27
IFNGR1
IL17F
IRAK4
IFNAR2
IL17RA
TYK2
IL17RC
IRF7
NCF2
GATA2
FERMT3
RAC2
CSF3R
ELANE
SP110
WIPF1
POLE
ORAI1
ATM
HELLS
NBN
EPG5
STAT5B
STIM1
TERT
Biliary−AdenoCA
Bladder−TCC
Breast−AdenoCA
Breast−LobularCA
Cervix−SCC
CNS−GBM
CNS−Medullo
CNS−Oligo
CNS−PiloAstro
ColoRect−AdenoCA
Eso−AdenoCA
Head−SCC
Kidney−ChRCC
Kidney−RCC
Liver−HCC
Lung−AdenoCA
Lung−SCC
Lymph−BNHL
Lymph−CLL
Myeloid−MPN
Ovary−AdenoCA
Panc−AdenoCA
Panc−Endocrine
Skin−Melanoma
Stomach−AdenoCA
Thy−AdenoCA
Uterus−AdenoCA
No.Carrier
0
2
4
6
8
Significant
FDR>0.1
FDR < 0.1
−log(FDR)
20
15
10
5
0

Table 2-2

Histology	Gene	PPV Prevalence	*P*-value	FDR
Bladder-TCC	*TNFRSF13B*	8.70%	1.01E-07	1.40E-05
Lung-AdenoCA	*ATM*	8.11%	0.001607895	0.066654538
Breast-LobularCA	*FERMT3*	7.69%	1.29E-13	4.19E-11
Cervix-SCC	*TYK2*	5.56%	2.95E-08	4.72E-06
Cervix-SCC	*TNFRSF13B*	5.56%	3.18E-05	0.002413713
CNS-Oligo	*ATM*	5.56%	6.25E-16	2.85E-13
Uterus-AdenoCA	*MEFV*	4.55%	9.95E-06	0.000934283
CNS-PiloAstro	*ATM*	4.49%	0.000271845	0.015223307
Bladder-TCC	*IRF7*	4.35%	7.85E-10	1.67E-07
Bladder-TCC	*IL17RC*	4.35%	2.75E-07	3.66E-05
Bladder-TCC	*ATM*	4.35%	0.001091647	0.049964896
Myeloid-MPN	*DOCK8*	3.85%	2.09E-06	0.000247526
Myeloid-MPN	*POLE*	3.85%	0.002503399	0.097449371
Stomach-AdenoCA	*ATM*	2.94%	0.000426883	0.022710185
Biliary-AdenoCA	*MEFV*	2.94%	0.000198147	0.011294361
Biliary-AdenoCA	*IRAK4*	2.94%	0.000970871	0.046254018
Biliary-AdenoCA	*NBN*	2.94%	0.000445182	0.0232807
Biliary-AdenoCA	*ITK*	2.94%	1.03E-08	1.74E-06
Lymph-BNHL	*TNFRSF13B*	2.80%	0.001787898	0.072240139
Lung-AdenoCA	*IRF7*	2.70%	4.45E-06	0.000473124

Lung-AdenoCA	*MS4A1*	2.70%	0.001325167	0.057119366
Lung-AdenoCA	*GATA2*	2.70%	4.19E-16	2.23E-13
CNS-GBM	*TRNT1*	2.56%	2.24E-05	0.001886752
CNS-GBM	*WIPF1*	2.56%	0.000513326	0.026008518
Panc-Endocrine	*CD8A*	2.47%	0.000452194	0.0232807
Kidney-ChRCC	*STAT5B*	2.33%	0.000118396	0.007558378
Kidney-ChRCC	*BLNK*	2.33%	0.002107285	0.08408067
Kidney-ChRCC	*CD3E*	2.33%	2.58E-05	0.002112589
Kidney-ChRCC	*CD247*	2.33%	0.000128486	0.007748592
Uterus-AdenoCA	*ELANE*	2.27%	9.40E-05	0.006252993
CNS-PiloAstro	*TYK2*	2.25%	0.000150757	0.008911435
Lung-SCC	*RELB*	2.13%	0.000128658	0.007748592
Thy-AdenoCA	*IFNAR2*	2.08%	4.48E-05	0.003175047
Thy-AdenoCA	*IL17F*	2.08%	4.55E-14	1.82E-11
Thy-AdenoCA	*CD3G*	2.08%	0.001195804	0.053013966
Thy-AdenoCA	*CR2*	2.08%	8.51E-06	0.000848979
Thy-AdenoCA	*NHEJ1*	2.08%	1.10E-06	0.000134487
ColoRect-AdenoCA	*DCLRE1C*	1.92%	2.97E-09	5.58E-07
ColoRect-AdenoCA	*IL7R*	1.92%	0.00053281	0.026573918
ColoRect-AdenoCA	*CSF3R*	1.92%	9.38E-06	0.000906992
ColoRect-AdenoCA	*ADA*	1.92%	9.40E-08	1.36E-05
Head-SCC	*CD27*	1.79%	4.12E-35	4.38E-32
Head-SCC	*CTPS1*	1.79%	0.00134209	0.057119366

Head-SCC	*ORAI1*	1.79%	3.04E-05	0.002413713
Head-SCC	*SP110*	1.79%	9.61E-19	6.14E-16
Stomach-AdenoCA	*STIM1*	1.47%	1.36E-05	0.001241944
Stomach-AdenoCA	*IGLL1*	1.47%	0.000121142	0.007582037
Stomach-AdenoCA	*RAC2*	1.47%	8.21E-12	2.02E-09
Stomach-AdenoCA	*EPG5*	1.47%	0.001746323	0.071464917
Stomach-AdenoCA	*HELLS*	1.47%	3.81E-05	0.002764214
Stomach-AdenoCA	*GATA2*	1.47%	1.36E-09	2.72E-07
Stomach-AdenoCA	*IFNGR1*	1.47%	0.000311018	0.017116722
Panc-Endocrine	*IRF7*	1.23%	0.00011124	0.007246515
Panc-Endocrine	*TERT*	1.23%	0.000665792	0.032200112
Panc-Endocrine	*CD81*	1.23%	6.69E-05	0.004541857
Panc-Endocrine	*JAK3*	1.23%	3.77E-09	6.69E-07
Panc-Endocrine	*TAP2*	1.23%	0.000423051	0.022710185
Panc-Endocrine	*TAPBP*	1.23%	0.001095721	0.049964896
Panc-Endocrine	*TNFRSF13C*	1.23%	5.38E-05	0.003733931
CNS-PiloAstro	*LCK*	1.12%	3.68E-05	0.002734566
Lymph-CLL	*CARD11*	1.11%	0.000574474	0.028211096
Lymph-CLL	*RAB27A*	1.11%	6.22E-06	0.000640147
Eso-AdenoCA	*FOXN1*	1.03%	0.000154431	0.008962622
Eso-AdenoCA	*UNC119*	1.03%	0.001176755	0.052904256
Eso-AdenoCA	*FERMT3*	1.03%	3.15E-05	0.002413713
Breast-AdenoCA	*IRF7*	1.03%	2.25E-05	0.001886752

Breast-AdenoCA	*IFNAR2*	1.03%	0.002452815	0.096659073
Skin-Melanoma	*IL17RA*	0.93%	5.39E-140	1.72E-136
Ovary-AdenoCA	*TERT*	0.91%	3.26E-06	0.000358403
Ovary-AdenoCA	*TYK2*	0.91%	6.48E-07	8.28E-05
Ovary-AdenoCA	*RELB*	0.91%	0.001579269	0.066329296
Ovary-AdenoCA	*DOCK2*	0.91%	3.87E-46	6.18E-43
Panc-AdenoCA	*AICDA*	0.86%	1.51E-05	0.001336515
CNS-Medullo	*CD27*	0.71%	9.05E-11	2.06E-08
CNS-Medullo	*ITK*	0.71%	0.00102491	0.048110478
CNS-Medullo	*NCF2*	0.71%	6.78E-20	5.41E-17
Kidney-RCC	*ELANE*	0.70%	7.06E-08	1.07E-05
Kidney-RCC	*LYST*	0.70%	0.001266786	0.055391501
Breast-AdenoCA	*CR2*	0.51%	4.90E-12	1.30E-09
Panc-AdenoCA	*IGLL1*	0.43%	3.24E-06	0.000358403
Liver-HCC	*DCLRE1C*	0.32%	1.31E-13	4.19E-11
Liver-HCC	*UNG*	0.32%	1.63E-12	4.72E-10

Figure 2-9-a

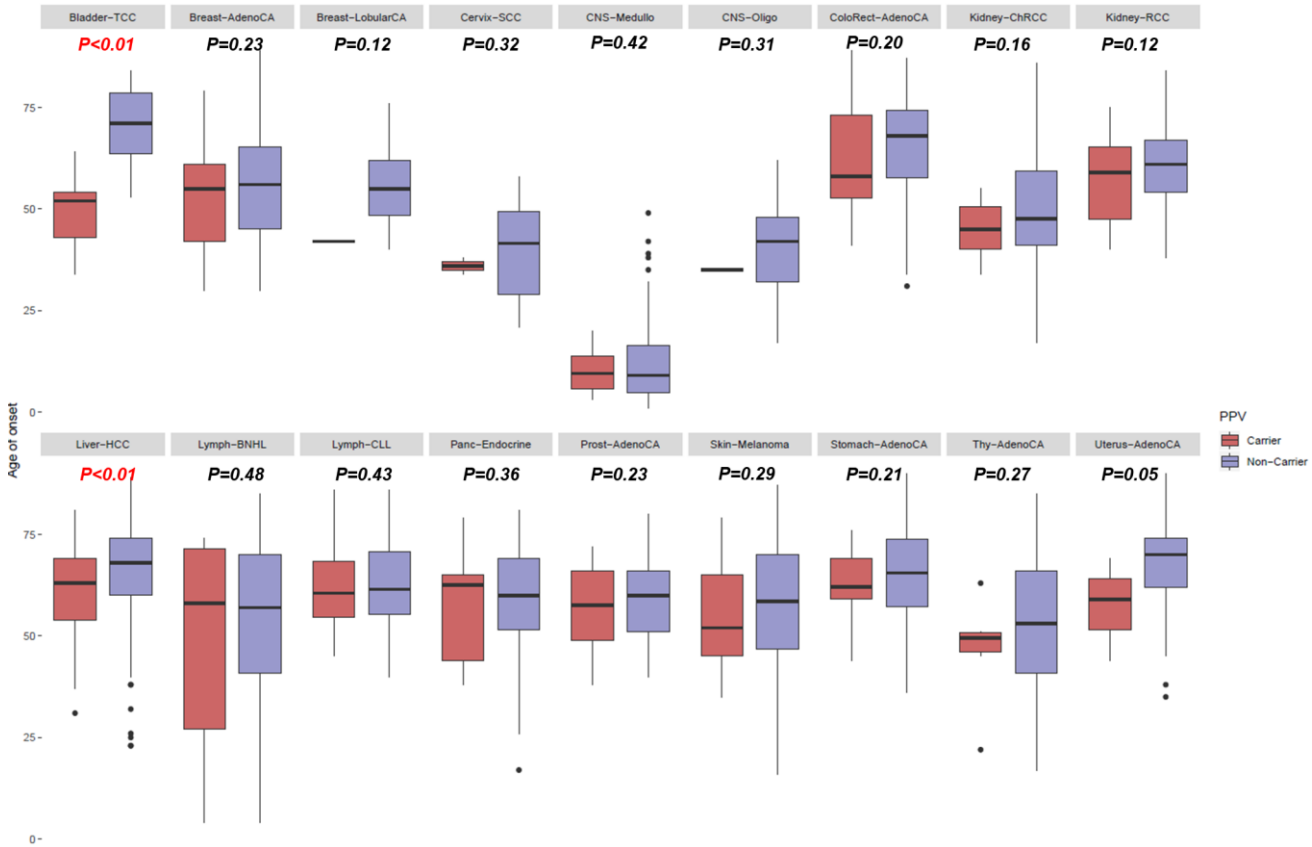

Bladder−TCC
Breast−AdenoCA
Breast−LobularCA
Cervix−SCC
CNS−Medullo
CNS−Oligo
ColoRect−AdenoCA
Kidney−ChRCC
Kidney−RCC
P<0.01
P=0.23
P=0.12
P=0.32
P=0.42
P=0.31
P=0.20
P=0.16
P=0.12
Liver−HCC
Lymph−BNHL
Lymph−CLL
Panc−Endocrine
Prost−AdenoCA
Skin−Melanoma
Stomach−AdenoCA
Thy−AdenoCA
Uterus−AdenoCA
P<0.01
P=0.48
P=0.43
P=0.36
P=0.23
P=0.29
P=0.21
P=0.27
P=0.05
Age of onset
75
50
25
0
PPV
Carrier
Non−Carrier

Figure 2-9-b

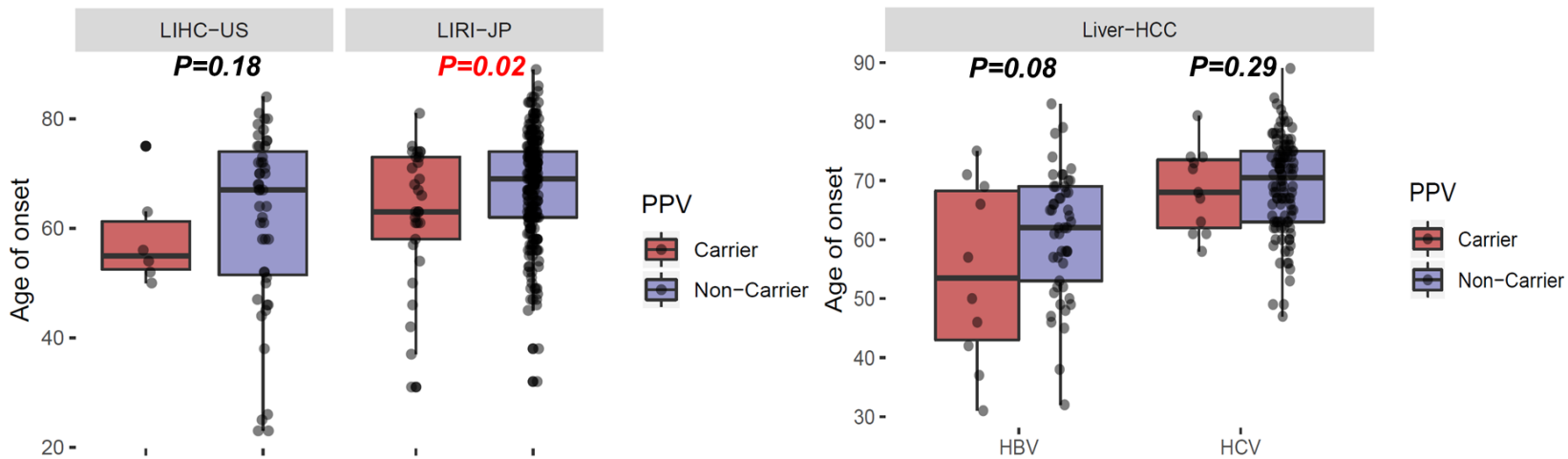

LIHC−US
LIRI−JP
P=0.18
P=0.02
Age of onset
80
60
40
20
PPV
Carrier
Non−Carrier
Liver−HCC
P=0.08
P=0.29
90
80
70
60
50
40
30
HBV
HCV
PPV
Carrier
Non−Carrier

Figure 2-10-a

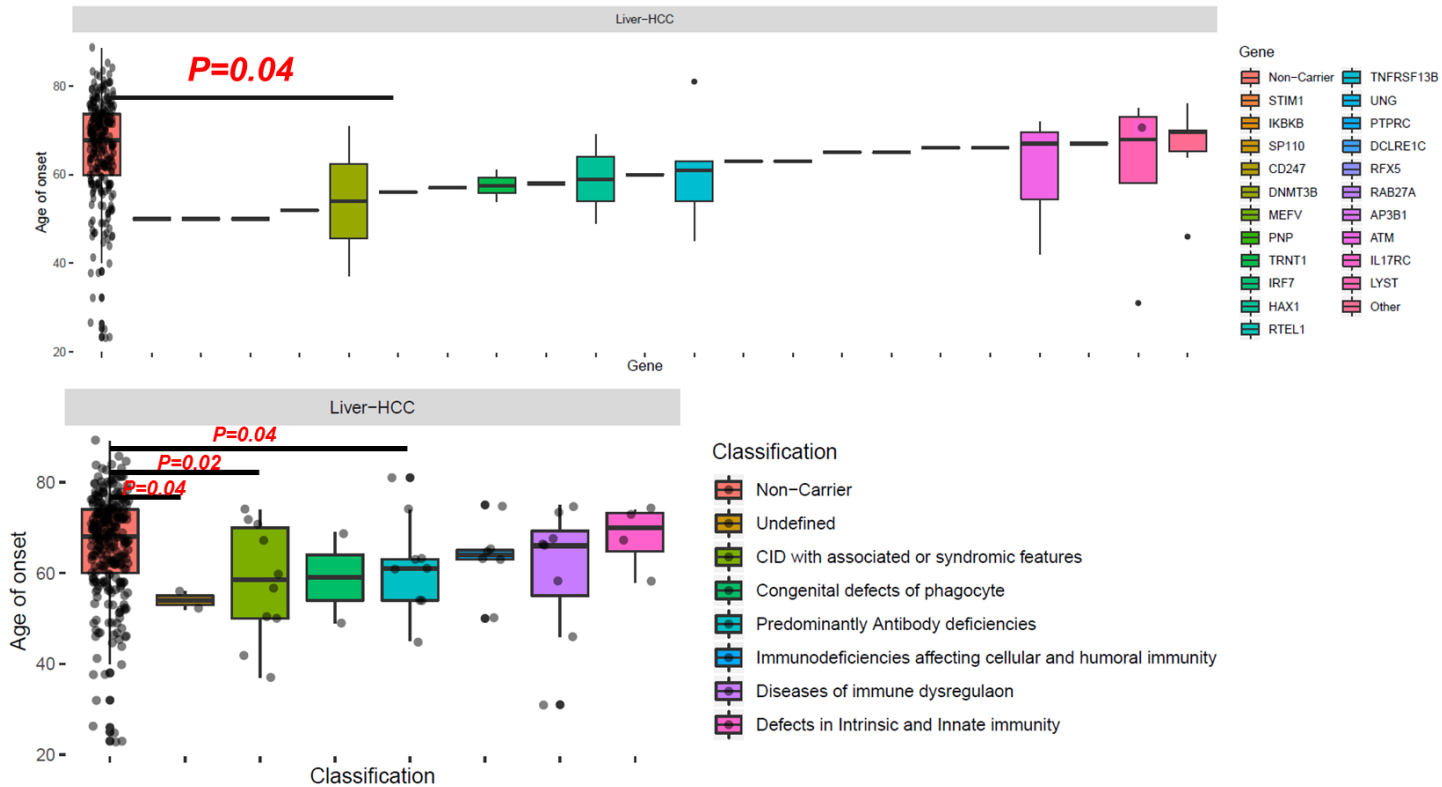
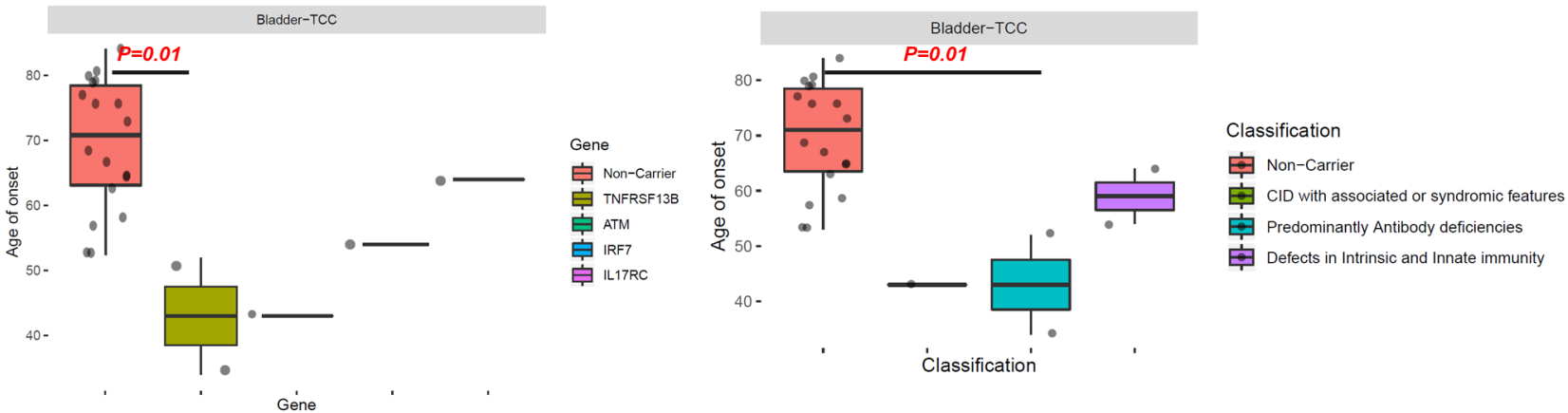

Liver−HCC
P=0.04
Age of onset
80
60
40
20
Gene
Gene
Non−Carrier
STIM1
IKBKB
SP110
CD247
DNMT3B
MEFV
PNP
TRNT1
IRF7
HAX1
RTEL1
TNFRSF13B
UNG
PTPRC
DCLRE1C
RFX5
RAB27A
AP3B1
ATM
IL17RC
LYST
Other
Liver−HCC
P=0.04
P=0.02
P=0.04
Age of onset
80
60
40
20
Classification
Classification
Non−Carrier
Undefined
CID with associated or syndromic features
Congenital defects of phagocyte
Predominantly Antibody deficiencies
Immunodeficiencies affecting cellular and humoral immunity
Diseases of immune dysregulaon
Defects in Intrinsic and Innate immunity

Figure 2-10-b


Bladder−TCC
P=0.01
Age of onset
80
70
60
50
40
Gene
Gene
Non−Carrier
TNFRSF13B
ATM
IRF7
IL17RC
Bladder−TCC
P=0.01
Age of onset
80
70
60
50
40
Classification
Classification
Non−Carrier
CID with associated or syndromic features
Predominantly Antibody deficiencies
Defects in Intrinsic and Innate immunity

Influence of DNA/RNA profiles and the immune cell milieu on PPV-bearing hepatocellular carcinoma and bladder cancer

Based on the promising results obtained from our genetic and clinical analyses, we focused on hepatocellular carcinoma and bladder cancer to determine whether PPV-bearing and PPV-free tumors show differentiating patterns in somatic mutations and gene expression. We therefore conducted RNA sequencing (RNA-seq) and identified a total of 604 genes that are differentially expressed in PPV-free and PPV-bearing hepatocellular carcinomas by DESeq2. These genes can be assigned to eight KEGG pathways, four of which are associated with cancer. In the case of bladder cancer, 1,433 genes were differentially expressed in PPV-free and PPV-bearing tumors, which are associated with 31 KEGG pathways (Figure 2-11-a,b). Somatic mutation profiles and signature analyses were also determined to further elucidate the basis for these observed differences in gene expression; however, no distinct signatures or mutation profiles were detected in either carcinoma (Figure 2-12).

Based on our finding that virus-associated hepatocellular carcinoma is more strongly associated with PPV-carrier status, we analyzed the immune cell milieu in this cancer subtype. Specifically, we measured the distribution of 22 types of immune cells infiltrating into the liver tissues using whole transcript sequencing in PPV carriers vs.

non-carriers. We found significant differences in two types of immune cells based on PPV-carrier status. That is, the infiltration of follicular helper T cells (Tfh) was significantly decreased in PPV carriers (P=0.04), whereas infiltration of regulatory T cells (Treg) was significantly increased in cancer tissues of PPV carriers (P=0.02) (Figure 2-13-a,b).

Family history of PPV-carrier status

To investigate the possibility that PPVs associated with PID are inherited from parents as germline factors, we examined family liver cancer history. We hypothesized that if PPVs inherited from parents affect cancer development, the parents of the cancer patients with PPVs should have a higher cancer risk, as compared to parents of the cancer patients without PPVs. Consistent with this, family history analysis revealed that parents of HCV-related liver cancer patients with PPVs had a 5.87-fold higher family history of cancer than parents of HCV liver cancer patients without PPVs (P=0.021; CI 95%: 1.05–29.13; Figure 2-14).

Figure Legends

Figure 2-11. Analysis of differentially expressed genes (DEGs) from liver and bladder cancer in PPV carriers vs. non-carriers

Hierarchical clustering of DEGs and gene set enrichment analysis from a) liver cancer and b) bladder cancer.

Figure 2-12. Mutation signature analysis

For each sample, the number of mutations corresponding to each signature is plotted. Orange bars indicate PPV carriers and green bars indicate non-carriers.

Figure 2-13. Infiltrating immune cell milieu in virus-related hepatocellular carcinoma from PPV carriers and non-carriers

Analysis of the immune cell milieu infiltrating into the liver tissue in subjects with virus-related hepatocellular carcinoma, as determined using RNA-seq: a) The percentage of whole immune cells in PPV carriers and non-carriers and b) A statistically significant difference in the numbers of infiltrating Treg and Tfh cells is observed in PPV carriers

vs. non-carriers.

Figure 2-14. Family cancer history of patients with virus-related hepatocellular carcinoma for PPV carriers and non-carriers

Odds ratios (ORs) were calculated at the 95% confidential interval level by the Fisher Exact test. Higher ORs indicate an increased cancer incidence in the family history of PPV carriers. Statistical significance is indicated in red (P <0.05). The top panel shows family history analysis for all cancer types, and the bottom panel shows liver cancer history.

Figure 2-11-a

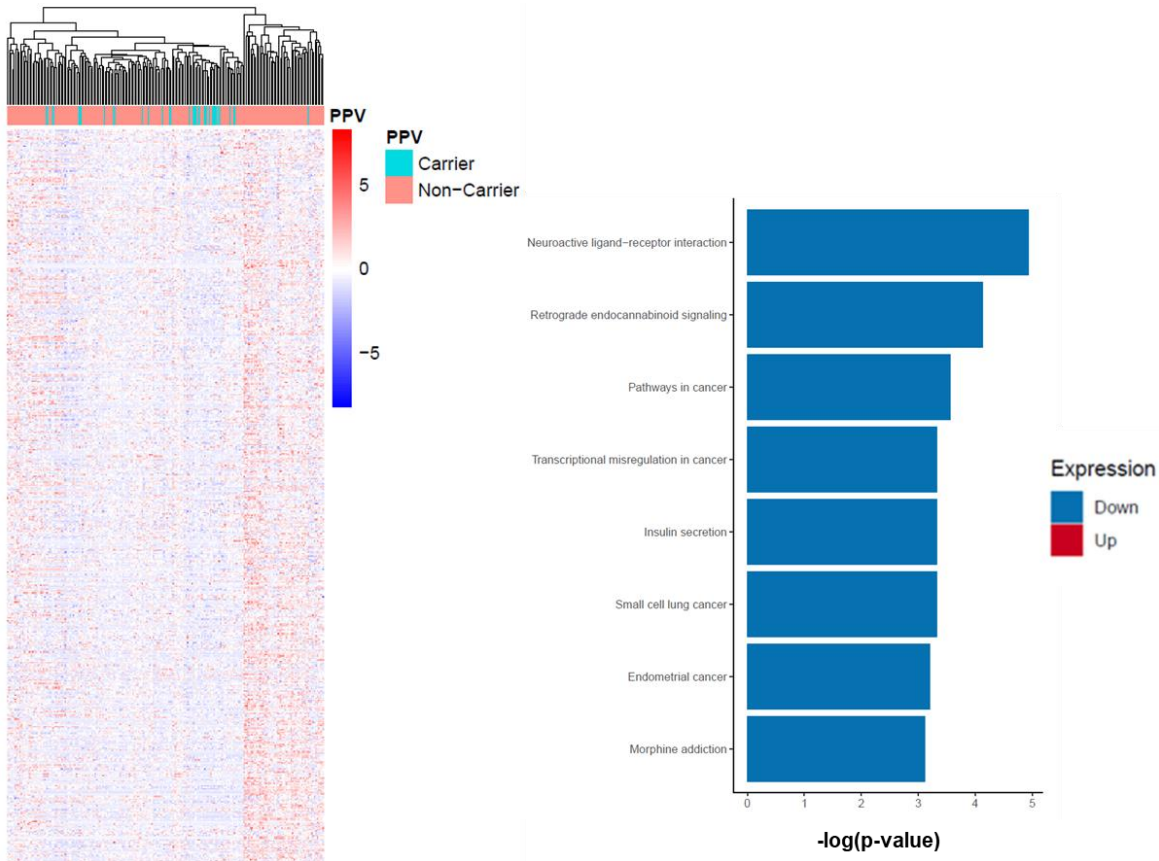

PPV
PPV
Carrier
Non−Carrier
5
0
−5
Neuroactive ligand−receptor interaction
Retrograde endocannabinoid signaling
Pathways in cancer
Transcriptional misregulation in cancer
Insulin secretion
Small cell lung cancer
Endometrial cancer
Morphine addiction
0
1
2
3
4
5
-log(p-value)
Expression
Down
Up

Figure 2-11-b

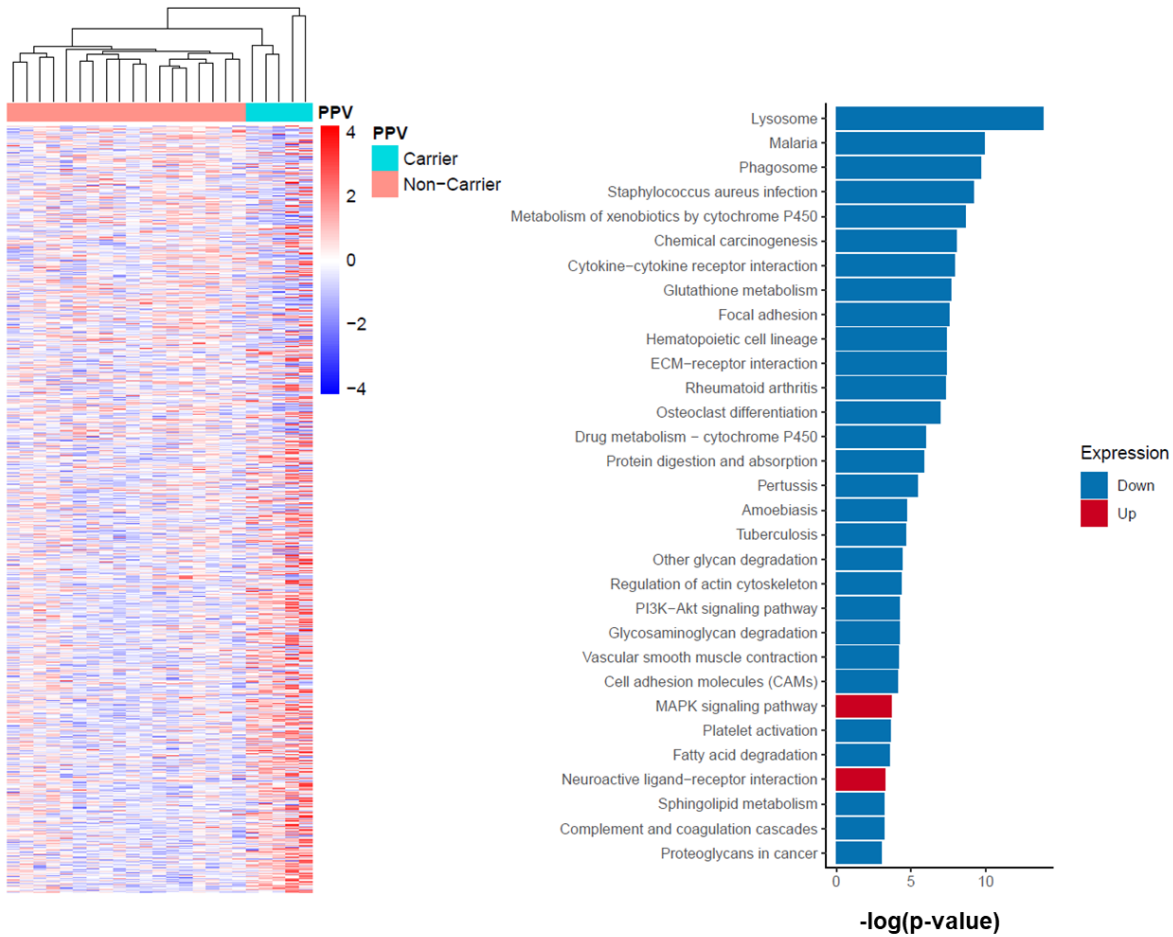

PPV
4
2
0
−2
−4
PPV
Carrier
Non−Carrier
Lysosome
Malaria
Phagosome
Staphylococcus aureus infection
Metabolism of xenobiotics by cytochrome P450
Chemical carcinogenesis
Cytokine−cytokine receptor interaction
Glutathione metabolism
Focal adhesion
Hematopoietic cell lineage
ECM−receptor interaction
Rheumatoid arthritis
Osteoclast differentiation
Drug metabolism − cytochrome P450
Protein digestion and absorption
Pertussis
Amoebiasis
Tuberculosis
Other glycan degradation
Regulation of actin cytoskeleton
PI3K−Akt signaling pathway
Glycosaminoglycan degradation
Vascular smooth muscle contraction
Cell adhesion molecules (CAMs)
MAPK signaling pathway
Platelet activation
Fatty acid degradation
Neuroactive ligand−receptor interaction
Sphingolipid metabolism
Complement and coagulation cascades
Proteoglycans in cancer
0
5
10
-log(p-value)
Expression
Down
Up

Figure 2-12

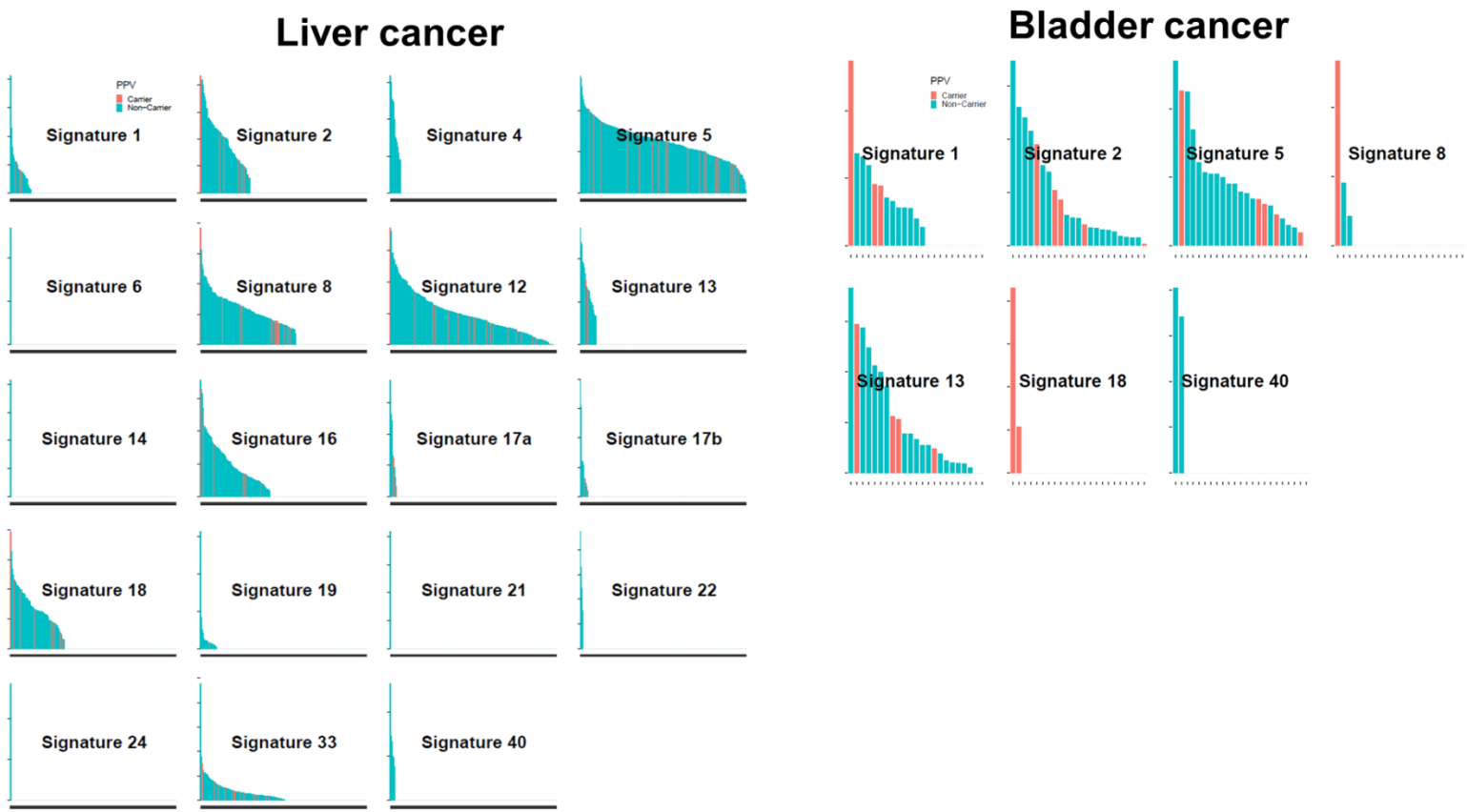

Liver cancer
PPV
Carrier
Non−Carrier
Signature 1
Signature 2
Signature 4
Signature 5
Signature 6
Signature 8
Signature 12
Signature 13
Signature 14
Signature 16
Signature 17a
Signature 17b
Signature 18
Signature 19
Signature 21
Signature 22
Signature 24
Signature 33
Signature 40
Bladder cancer
PPV
Carrier
Non−Carrier
Signature 1
Signature 2
Signature 5
Signature 8
Signature 13
Signature 18
Signature 40

Figure 2-13-a

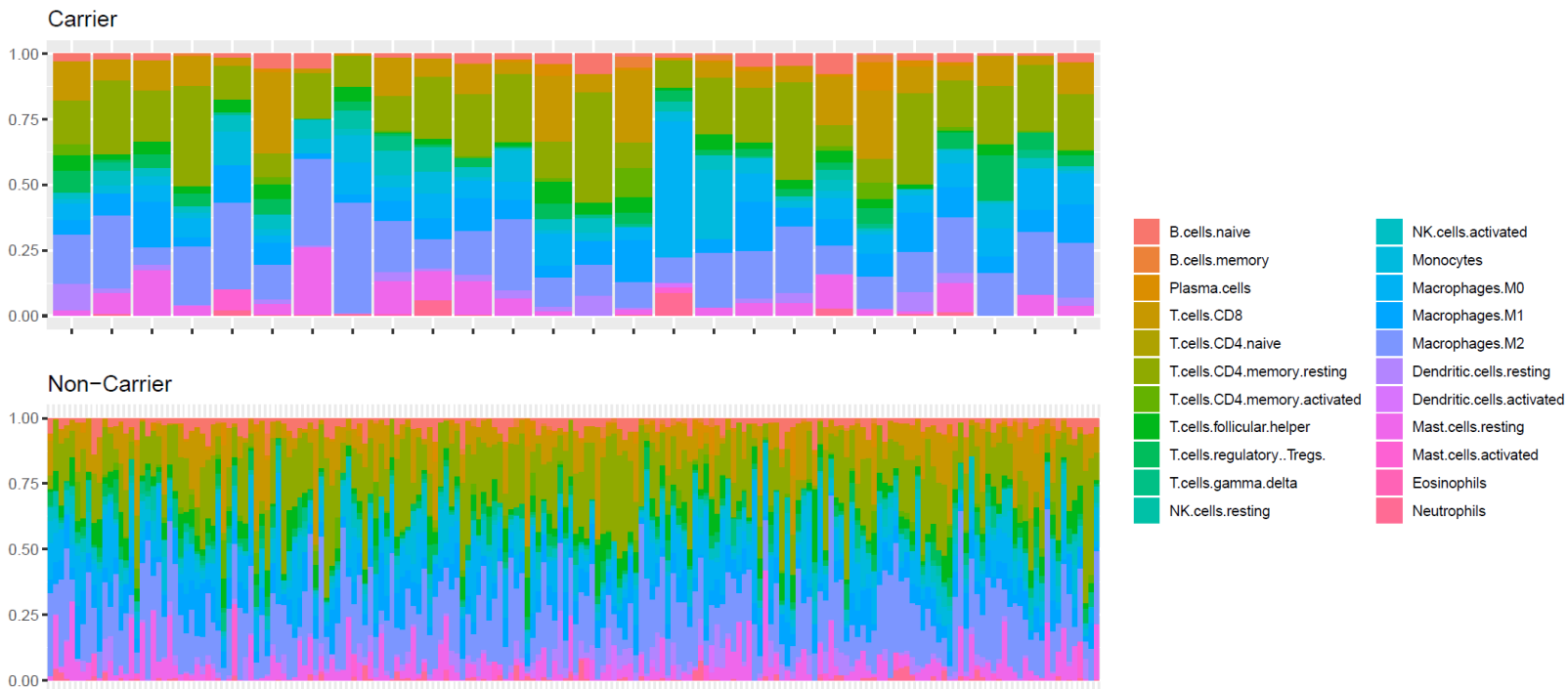

Carrier
1.00
0.75
0.50
0.25
0.00
Non−Carrier
1.00
0.75
0.50
0.25
0.00
B.cells.naive
B.cells.memory
Plasma.cells
T.cells.CD8
T.cells.CD4.naive
T.cells.CD4.memory.resting
T.cells.CD4.memory.activated
T.cells.follicular.helper
T.cells.regulatory..Tregs.
T.cells.gamma.delta
NK.cells.resting
NK.cells.activated
Monocytes
Macrophages.M0
Macrophages.M1
Macrophages.M2
Dendritic.cells.resting
Dendritic.cells.activated
Mast.cells.resting
Mast.cells.activated
Eosinophils
Neutrophils

Figure 2-13-b

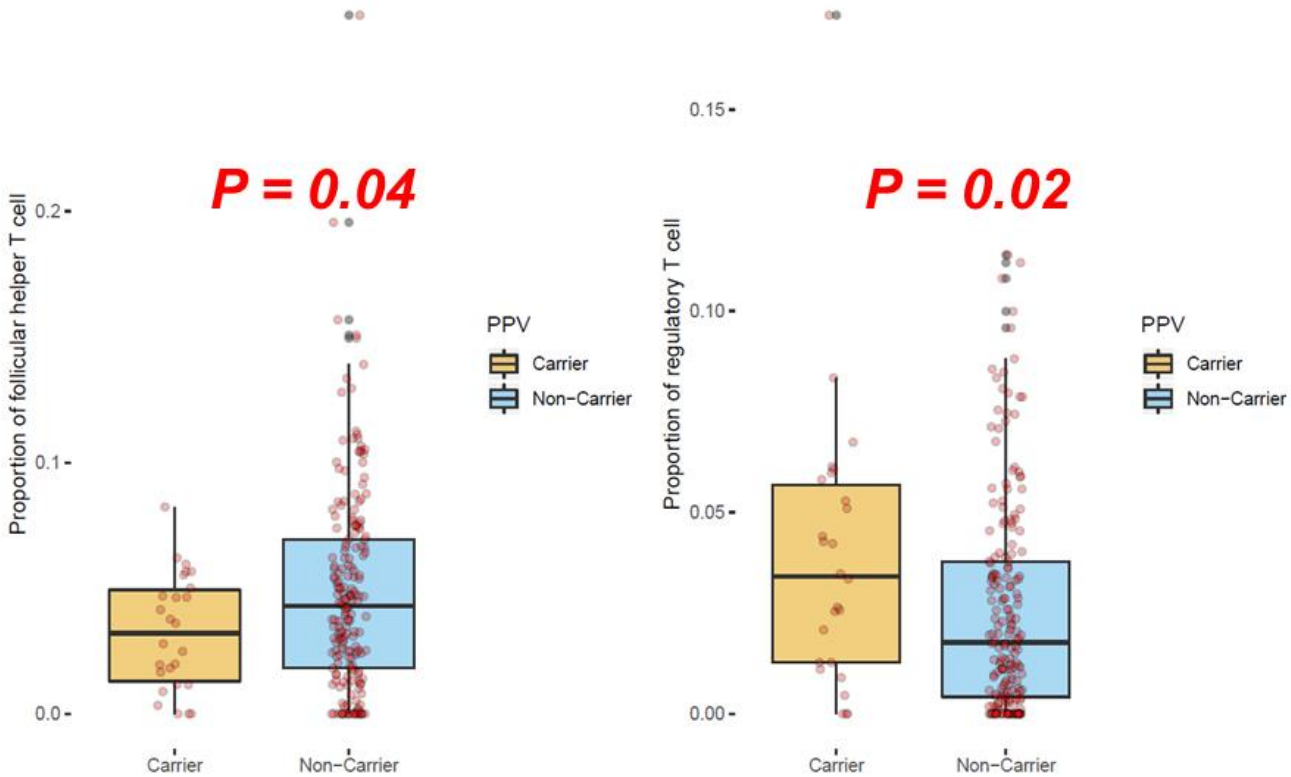
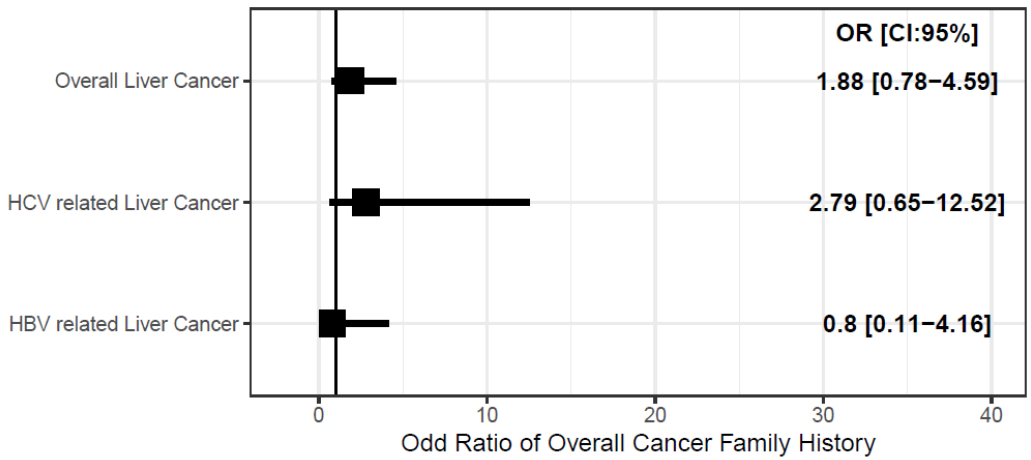
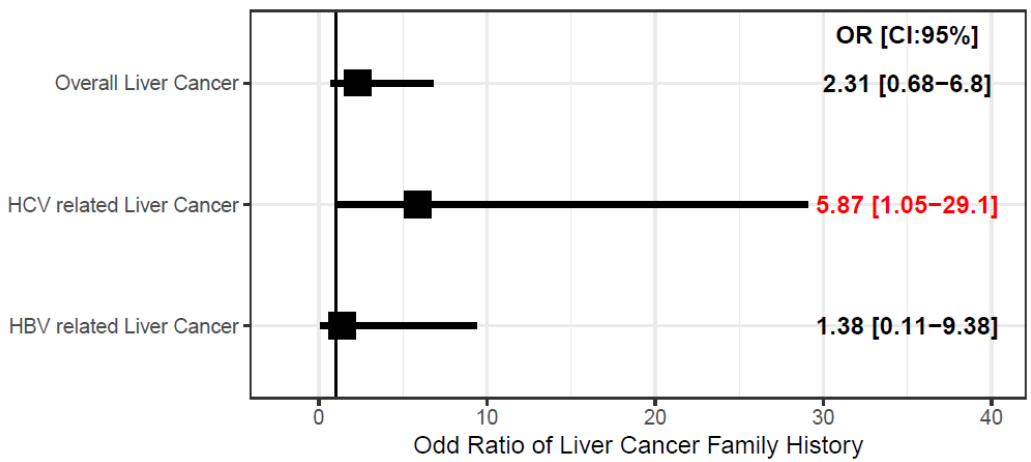

P = 0.04
Proportion of follicular helper T cell
0.2
0.1
0.0
Carrier
Non−Carrier
PPV
Carrier
Non−Carrier
P = 0.02
Proportion of regulatory T cell
0.15
0.10
0.05
0.00
Carrier
Non−Carrier
PPV
Carrier
Non−Carrier

Figure 2-14


OR [CI:95%]
Overall Liver Cancer
1.88 [0.78−4.59]
HCV related Liver Cancer
2.79 [0.65−12.52]
HBV related Liver Cancer
0.8 [0.11−4.16]
0
10
20
30
40
Odd Ratio of Overall Cancer Family History


OR [CI:95%]
Overall Liver Cancer
2.31 [0.68−6.8]
HCV related Liver Cancer
5.87 [1.05−29.1]
HBV related Liver Cancer
1.38 [0.11−9.38]
0
10
20
30
40
Odd Ratio of Liver Cancer Family History

Impact of gender on PPV-bearing hepatocellular carcinoma

To investigate the potential for gender bias in the immune surveillance of cancer cells, we performed DEG analysis segregated by gender for HBV-infected hepatocellular carcinoma patients, the most relevant disease in the above analysis. Hierarchical clustering was performed using the 163 DEGs identified previously during analysis to determine whether these were well-reflected in PPV-bearing cancer in both men and women. This analysis revealed that hierarchical clustering more accurately reflects PPV-carrier status in female HBV patients than in male HBV patients. In addition, gene set analysis identified two KEGG pathways (Mineral absorption, $P<0.001$, and p53 signaling pathway, $P=0.01$) that were significantly segregated by PPV-carrier status in women with HBV-related hepatocellular carcinoma, and interestingly, of these, the p53 pathway was down-regulated. Conversely, although 149 genes were found to be differentially expressed in men with HBV-related hepatocellular carcinoma, specific pathways associated with these genes were not detected (Figure 2-15-a,b). To verify whether the DEG result of the TP53 pathway was an error, three cases and six control groups were randomly selected and subjected to DEG analysis 30 times. As a result of the 30 permutations, the DEGs associated with the TP53 pathway were not found.

Figure Legend

Figure 2-15. DEG analysis by gender in Japanese hepatocellular carcinoma patients

a) Hierarchical clustering of RNA expression profiles in female patients with HBV-related liver cancer is shown in the left panel, and the TP53 signaling map from KEGG pathway analysis is shown on the right. Green shading indicates down-regulated gene expression. b) Hierarchical clustering of RNA expression profiles in male patients with HBV-related liver cancer is shown.

Figure 2-15-a

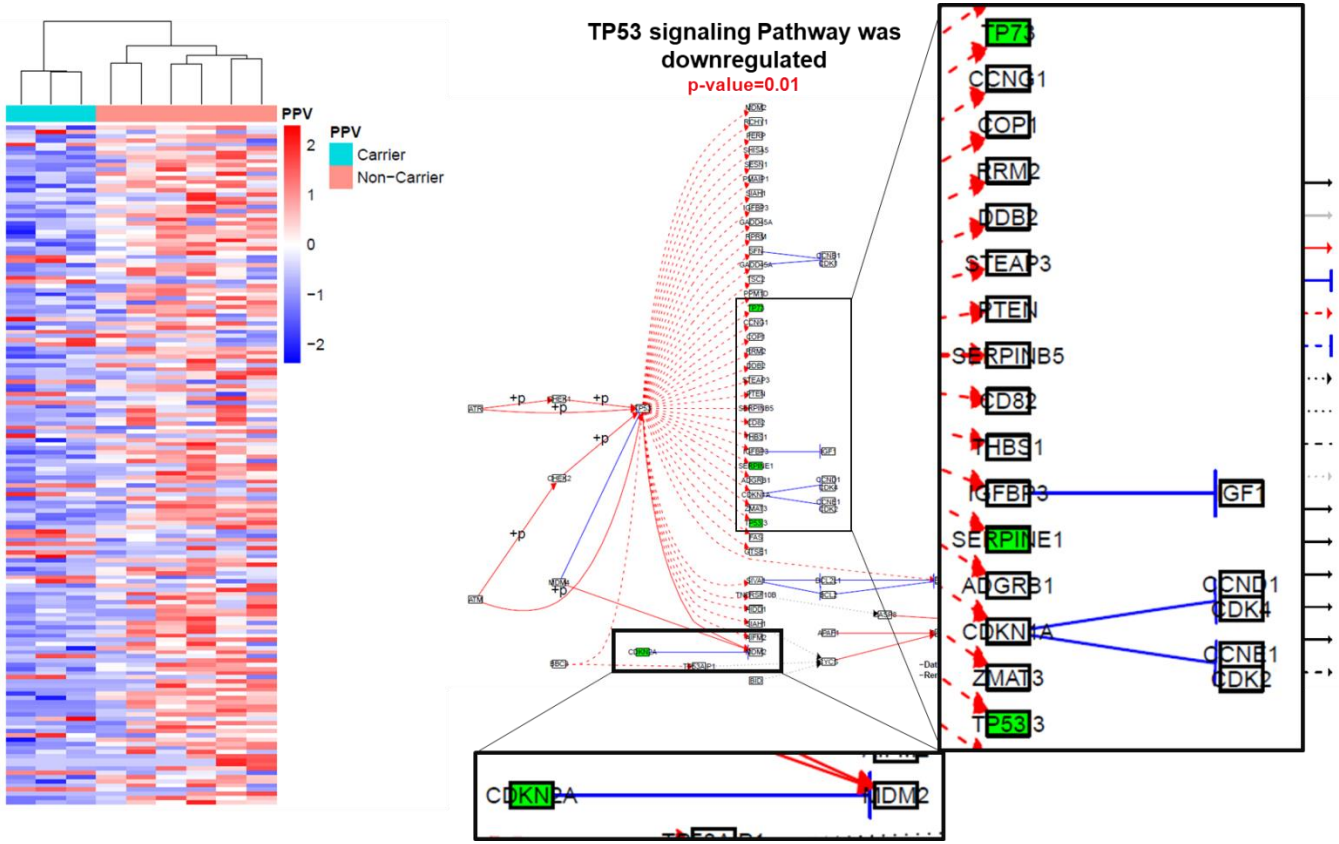
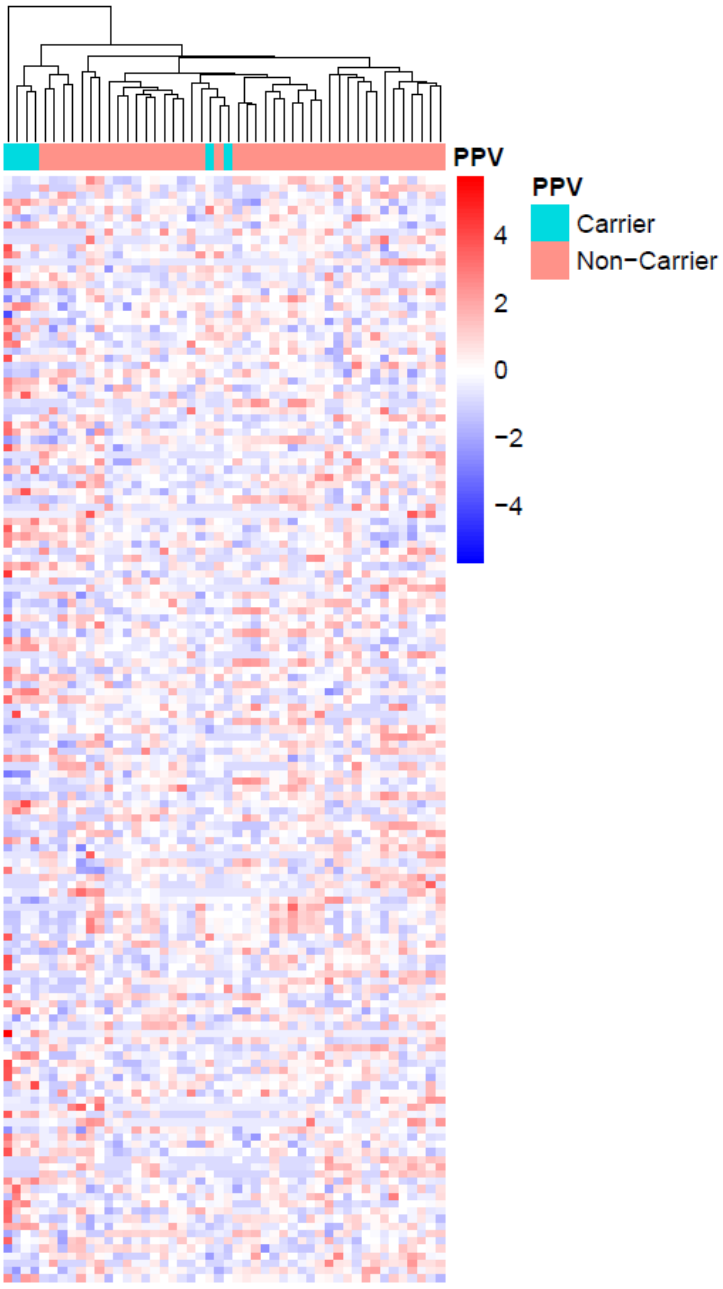

PPV
2
1
0
−1
−2
PPV
Carrier
Non−Carrier
TP53 signaling Pathway was downregulated
p-value=0.01
TP73
CCNG1
COPS
RRM2
DDB2
STEAP3
PTEN
SERPINB5
CD82
THBS1
IGFBP3
GF
SERPINE1
ADGRB1
CCND1
CDK4
CDKN1A
CCNB1
CDK2
ZMAT3
TP53I3
CDKN2A
MDM2

Figure 2-15-b



External validation of the clinical significance of PPVs carrier status using targeted sequencing

To verify the clinical significance of the 114 PID-related genes, we measured PPV status in an external cohort of 1,667 Korean patients with four types of cancer. Specifically, we analyzed the association between the age of cancer onset and PPV-carrier status in subjects with bladder cancer or liver cancer. For this analysis, we considered mutations found in >1% of the cohort as Korean-specific SNPs, and excluded these mutations from the test.

The hepatocellular carcinoma cohort consisted of a total of 463 patients: 401 with HBV-associated liver cancer (n=322 male, n=79 female), 61 with HCV-associated liver cancer (n=44 male, n=17 female), and one subject with both HBV and HCV. We found that the age of onset for PPV carriers in the overall cohort tended to be lower, as compared to non-carriers (median age: 57 for PPV carriers vs. 59 for PPV non-carriers; P=0.05). PPV status also appears to have a greater effect on age of onset in patients with HBV-related hepatocellular carcinoma (median age: 57 for PPV carriers vs. 59 for PPV non-carriers; P=0.05), than in patients with HCV-related hepatocellular carcinoma (median age: 62 for PPV carriers vs. 67 for PPV non-carriers; P=0.35, Figure 2-17-a).

Analysis of the association between age of onset and PID gene

immune pathology classification revealed that similar to our previous findings from the Pan-Cancer cohort, decreased age of onset was significantly associated with PPVs in “Predominantly antibody deficiency” genes (P=0.01, Figure 2-17-b). In female patients in particular, cancer development in PPV carriers was 12 years earlier than non-carriers, implying that immune surveillance has a gender-specific impact (median age: 52 for carriers vs. 64 for non-carriers; P<0.01, Figure 2-17-c). Further, the age of cancer onset in women with HBV-related hepatocellular carcinoma was also significantly decreased (median age: 51 for carriers vs. 62 for non-carriers; P=0.01, Figure 2-17-d). Notably, although age of onset did not show a statistically significant association with PPV-carrier status in the LIRI-JP subset of the Pan-Cancer cohort, the trends were consistent for female subjects (mean onset age for women with HBV-associated liver cancer: 57 vs. 59, and mean onset age for women with HCV-associated liver cancer: 67 vs. 70; Figure 2-16).

When PID genes were classified by immune pathology, we found that PPV carriers in “CID with associated or syndromic features” and “Predominantly antibody deficiencies” categories show reduced age of cancer onset, consistent with our previous results. Among these classifications, PPVs in the *ATM* gene in “CID with associated or syndromic features” and the *TNFRSF12B* gene in “Predominantly

antibody deficiencies" were recurrently found in the ICGC cohort and external cohort.

Figure Legends

Figure 2-16. Age of onset for HBV/HCV-related liver cancer by gender

The points represent the age of cancer onset for each individual. The lines connect the mean age of onset in PPV carriers and PPV non-carriers and are colored by gender (red = female and blue = male).

Figure 2-17. Association between the age of cancer onset and PPV-carrier status in the Korean liver validation cohort

a) The overall age of onset in PPV carriers and non-carriers is graphed on the left. Graphs on the right show age of onset for HBV-related and HCV-related liver cancer in PPV carriers and non-carriers. b) Age of onset in PPV carriers in each of the seven groups of PID genes, classified by immune pathology is shown. c) The graph on the left shows the age of onset in male and female PPV carriers and non-carriers. Graphs on the right show age of onset in PPV carriers, grouped by gender, in each of the seven groups of PID genes, classified by immune pathology. d) Age of onset in PPV carriers and non-carriers grouped by gender and disease etiology (e.g., HBV or HCV) is shown.

Figure 2-16

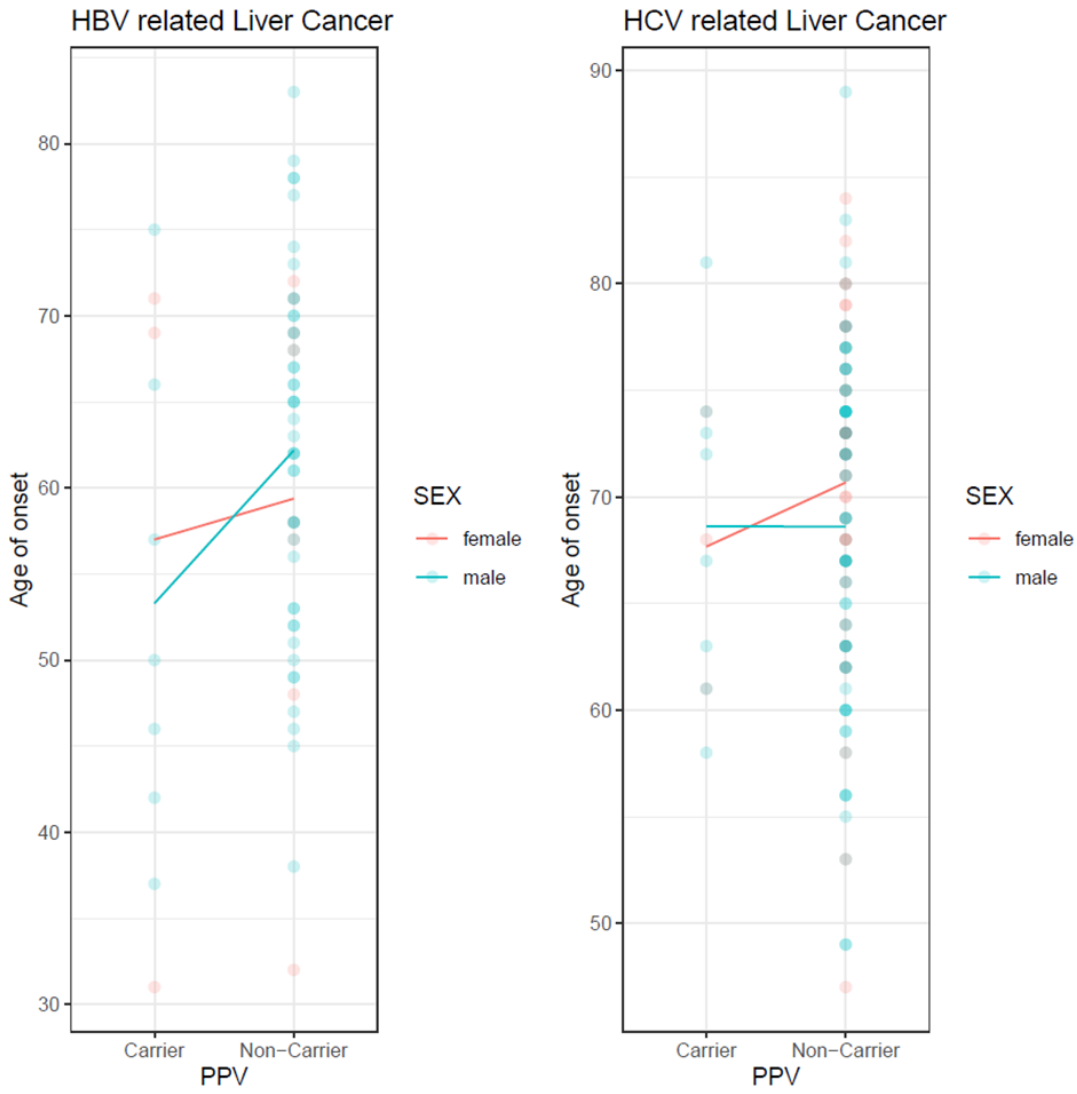

HBV related Liver Cancer
Age of onset
80
70
60
50
40
30
Carrier
Non−Carrier
PPV
SEX
female
male
HCV related Liver Cancer
Age of onset
90
80
70
60
50
Carrier
Non−Carrier
PPV
SEX
female
male

Figure 2-17-a

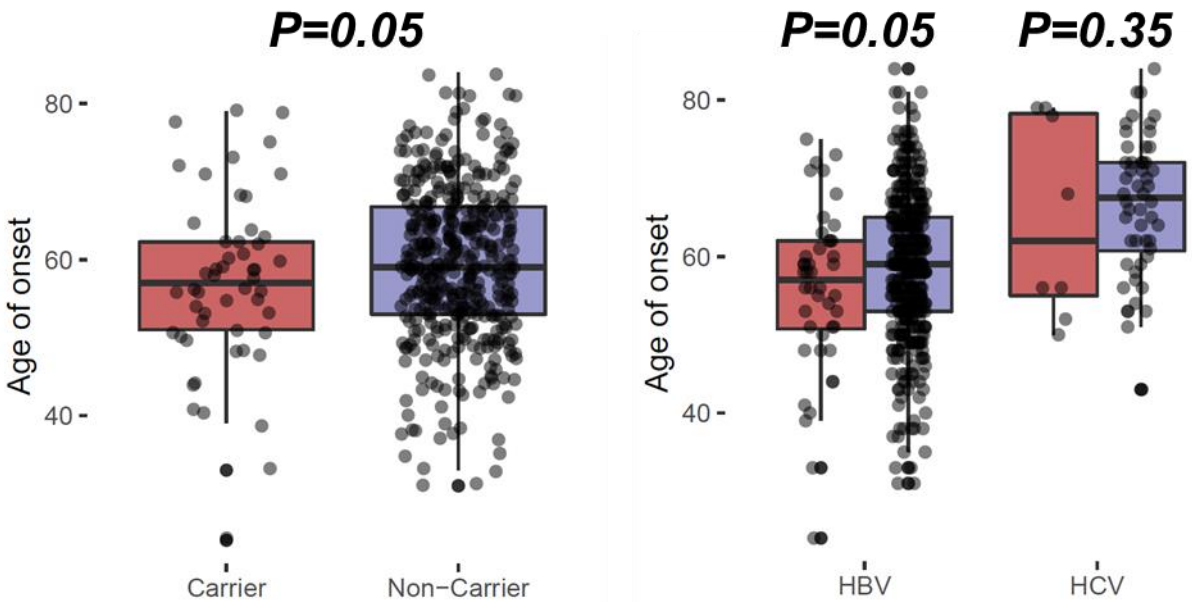

P=0.05
80
60
40
Age of onset
Carrier
Non−Carrier
P=0.05
P=0.35
80
60
40
Age of onset
HBV
HCV

Figure 2-17-b

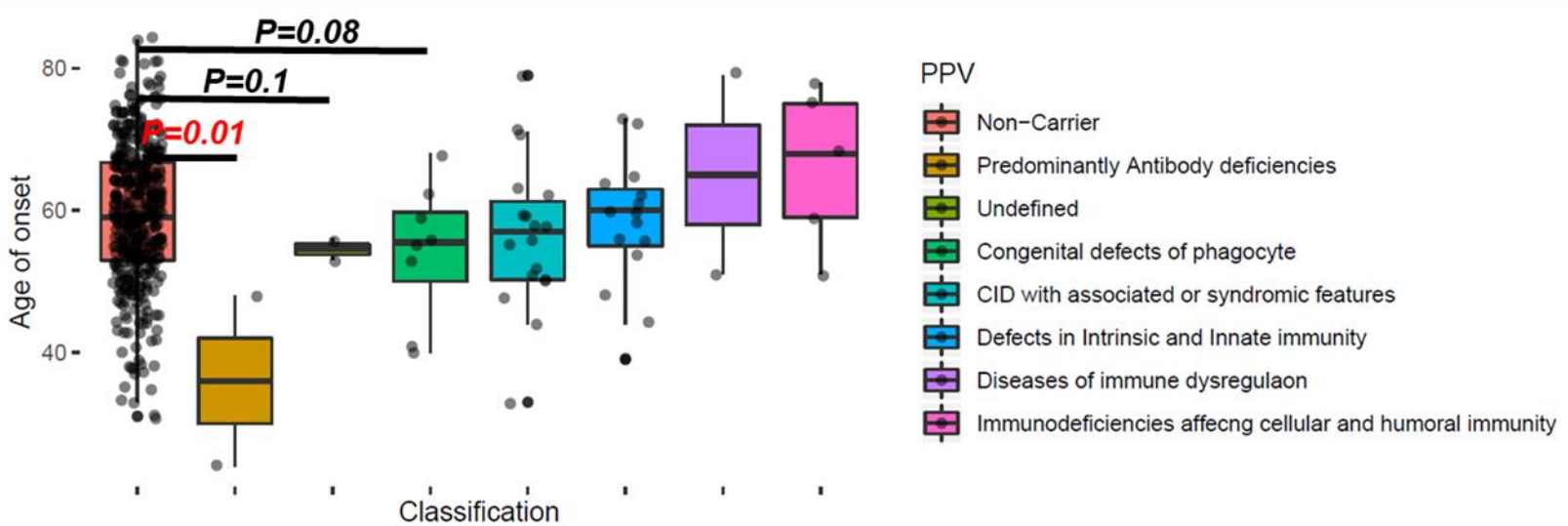
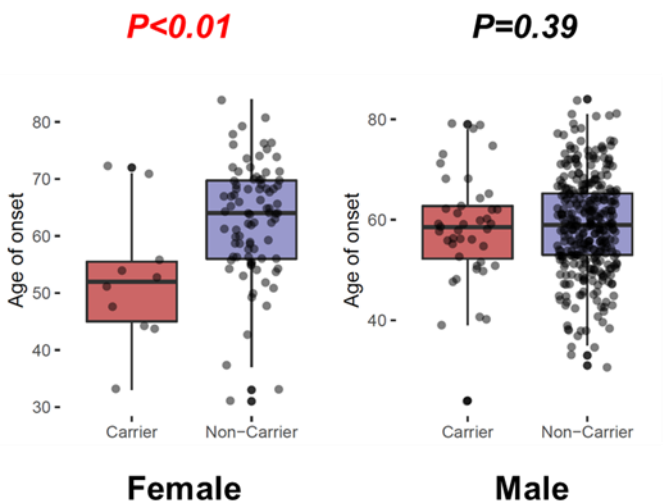

P=0.08
P=0.1
P=0.01
80
60
40
Age of onset
Classification
PPV
Non−Carrier
Predominantly Antibody deficiencies
Undefined
Congenital defects of phagocyte
CID with associated or syndromic features
Defects in Intrinsic and Innate immunity
Diseases of immune dysregulaon
Immunodeficiencies affecng cellular and humoral immunity

Figure 2-17-c


P<0.01
P=0.39
Age of onset
80
70
60
50
40
30
Carrier
Non−Carrier
Female
Male

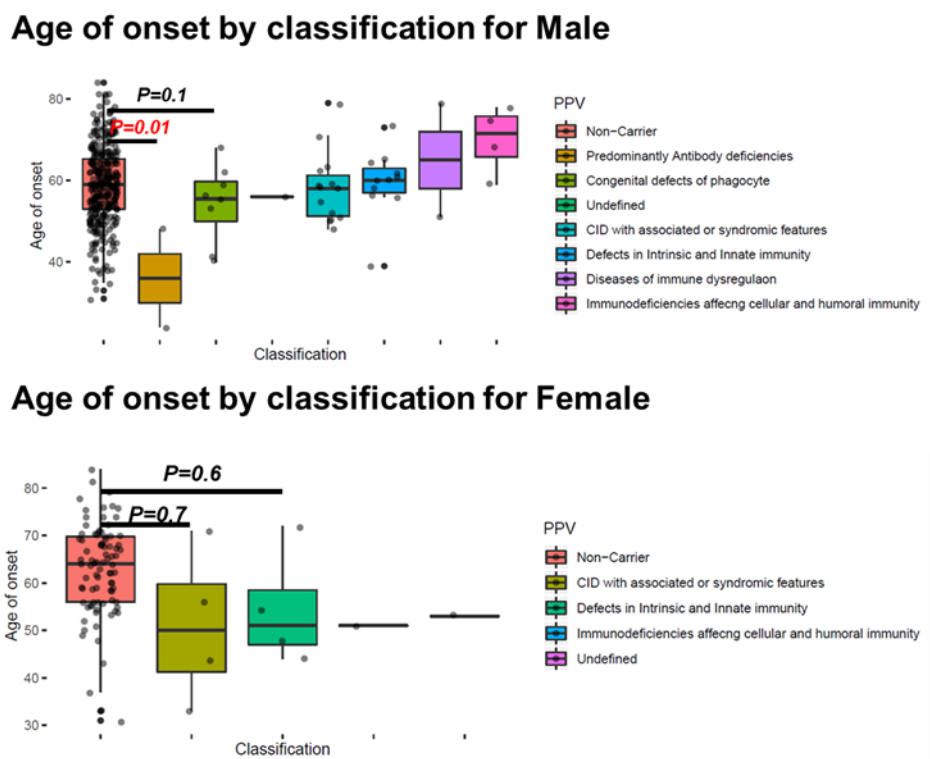
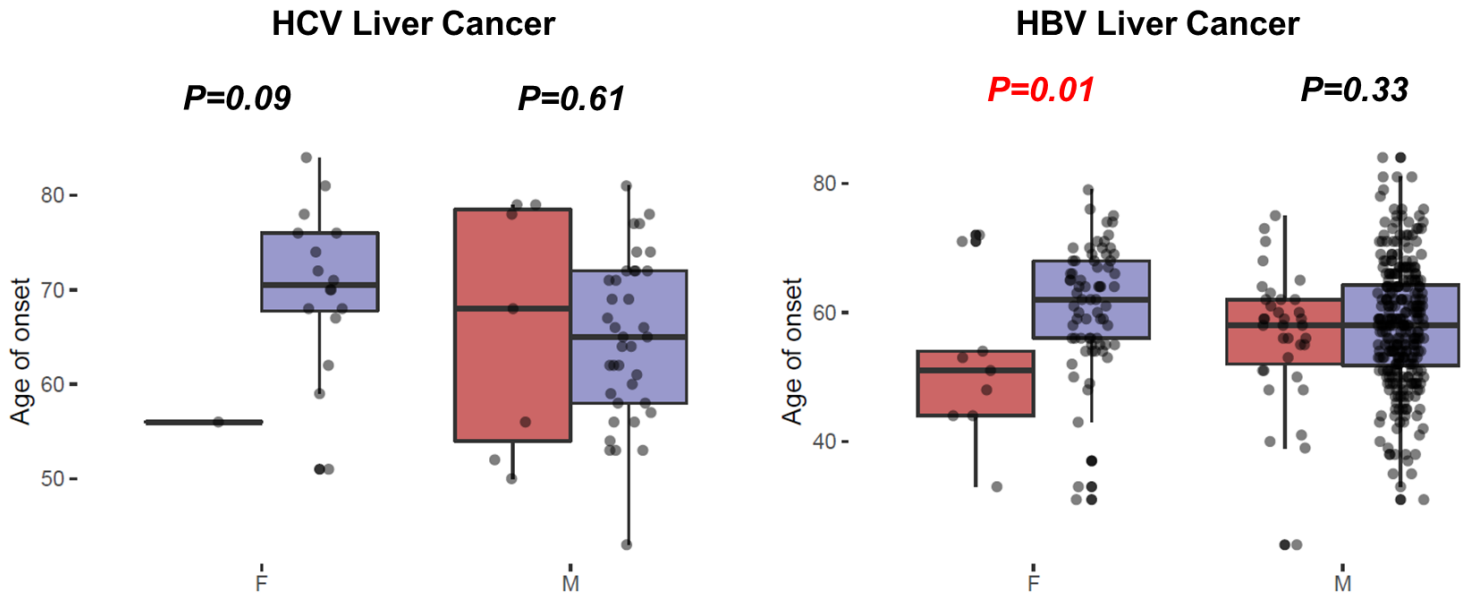

Age of onset by classification for Male
P=0.1
P=0.01
Age of onset
Classification
PPV
Non−Carrier
Predominantly Antibody deficiencies
Congenital defects of phagocyte
Undefined
CID with associated or syndromic features
Defects in Intrinsic and Innate immunity
Diseases of immune dysregulaon
Immunodeficiencies affecng cellular and humoral immunity
Age of onset by classification for Female
P=0.6
P=0.7
Age of onset
Classification
PPV
Non−Carrier
CID with associated or syndromic features
Defects in Intrinsic and Innate immunity
Immunodeficiencies affecng cellular and humoral immunity
Undefined

Figure 2-17-d


HCV Liver Cancer
P=0.09
P=0.61
HBV Liver Cancer
P=0.01
P=0.33
Age of onset
80
70
60
50
40
F
M

DISCUSSION

The Mendelian genetic disease has long been known to occur in children; however, recent cases of adult-onset Mendelian genetic disease have also been reported, such as with adult Gaucher disease (Guimaraes et al., 2003). This suggests that mild carriers with inherited genetic mutations, who do not develop severe disease in childhood, can survive until adulthood with partially defective gene function. From an immunological perspective, in many cases, these mild carriers are expected to experience a decline of immune surveillance, which may later contribute to the development of cancer.

In this study, we screened a cancer patient cohort and a normal control cohort, both of which contained a population of individuals with undiagnosed PID, and found that even individuals who are not diagnosed with severe disease can harbor PID-associated germline mutations. Moreover, using these large-scale WGS datasets, we uncovered underlying associations between mild carriers of PID and increased cancer risk. Both liver and bladder cancer showed increased prevalence and decreased age of cancer onset in PPV carriers vs. non-carriers for 114 PID-associated genes, suggesting that immune surveillance plays an important role in cancer development of the liver and bladder.

The critical role of liver function in immune defense against various microbes and toxins is well-known. For example, the liver is required for immunological surveillance of the massive antigen load in the gastrointestinal tract, and it is involved in the systemic inflammatory response, induction of immunological tolerance, and destruction of activated lymphocytes. Thus, this organ plays essential roles in the first line immune defense against invading pathogens, the modulation of liver injury, and the recruitment of circulating lymphocytes (Racanelli and Rehermann, 2006). The role of the liver as an immunological organ is evidenced by the fact that approximately one-fourth of patients

with PID have some form of accompanying liver failure, with CID and common variable immunodeficiency (CVID) being the most common PIDs associated with liver disease (Rodrigues et al., 2004). CID is a complex immunodeficiency disease with various syndromes (here classified as "CID with associated or syndromic features"), and CVID is a condition associated with defects in overall antibody production (classified here with "Predominantly antibody deficiencies").

Our observed association between bladder cancer and immune system function can be explained by epidemiological observations in CVID Patients. Consistent with our findings, it was previously found that the incidence of bladder cancer in patients with CVID was around 4-fold higher than in those without CVID (Vajdic et al., 2010). Here, we found that CID and CVID were two of the most frequently found classifications and were the main determinants of clustering. In addition, liver cancer patients with PPVs in genes associated with CID and CVID developed cancer significantly earlier compared with wild-type liver cancer patients. Moreover, CID and CVID showed similar tendencies in our external validation cohort, strongly affirming our findings.

CID with associated or syndromic features includes nine pathogenic features: congenital thrombocytopenia, DNA repair defects, immuno-osseous dysplasia, thymic defects with additional congenital anomalies, hyper-IgE syndromes, dyskeratosis congenita (DKC), defects of vitamin B12 and folate metabolism, anhidrotic ectodermal dysplasia, and others (Bousfiha et al., 2018). Among these, most defective genes that have been identified in CID are involved in DNA repair, including *ATM*, which was frequently found in our analysis. This gene is known to function in DNA repair, but it is also associated with various immunodeficiency symptoms, such as telangiectasia, pulmonary infections, and decreased IgA, IgE, and IgG subclasses (Lumsden et al., 2004; McGrath-Morrow et al., 2010; McKinnon, 2004). In addition, an elevated prevalence of *ATM* germline mutations in cancer is known to increase the incidence of breast, gastric, and

colorectal cancer (Huang et al., 2015; Tavera-Tapia et al., 2017; Thompson et al., 2005). Loss of *ATM* also decreases levels of hepatocellular apoptosis and fibrosis in a high-fat diet-induced mouse model of non-alcoholic fatty liver disease, suggesting that this gene may play an important role in liver cancer (Daugherity et al., 2012). Based on these findings, we speculate that the mechanistic association between *ATM* and carcinogenesis detected in this study is due to the loss of its DNA repair function. However, due to the fact that *ATM* is also involved in various immunodeficiency diseases, it is possible that its effect on immune function also contributes to carcinogenic mechanisms.

Another notable classification detected in this study, "Predominantly antibody deficiencies," is composed of two immunological pathologies, including hypogammaglobulinemia and other antibody deficiencies (Bousfiha et al., 2018). A direct association between predisposing germline mutations in the *TNFRSF13B* gene and cancer development has not yet been established. However, *TNFRSF13B* deficiency is known to play a direct role in hypogammaglobulinemia, and the epidemiologic observation of elevated cancer development in patients with hypogammaglobulinemia has also been reported (Leone et al., 2018; Mohammadi et al., 2009).

We further found that the elevated cancer incidence associated with PID was significantly affected by gender for HBV-related liver cancer. This phenomenon is believed to arise from differences in immune surveillance due to gender bias in various immune-related genes that are linked to hormones or the X-chromosome (Ellegren and Parsch, 2007; Fish, 2008; Oertelt-Prigione, 2012). In addition to genetic factors, carcinogenic chemical exposures are also considered to be one of the main causes of this gender bias. Generally, men are more exposed to carcinogenic chemicals, such as alcohol and tobacco, than women. In addition to differences in the frequency of physical exposures, Kadekar et al. found that men had a higher risk of cancer development from exposure to carcinogenic chemicals (Kadekar et al., 2012). Thus, the effect of genetic

factors on cancer in men is blurred, suggesting a possible reason that women were found to be more likely to develop cancer due to genetic factors.

We also performed immune cell milieu analysis using RNA-seq data from viral-associated hepatocellular carcinoma and found that intra-tumor infiltration of follicular helper T (Tfh) cells and regulatory T (Treg) cells were altered in PPV-bearing vs. PPV-free tumors. Tfh cells represent a distinct subset of $CD4^+$ T cells that specialize in providing help to B lymphocytes and are involved in infectious diseases, allergy, and autoimmunity. A strong positive correlation has been detected between prognosis and Tfh cells for several cancers, including breast and colorectal cancer, thus providing strong evidence these cells play a role in cancer immunity (Bindea et al., 2013; Gu-Trantien et al., 2013). Treg cells, which are characterized by expression of the master regulatory transcription factor, *FOXP3*, are a highly immunosuppressive subset of $CD4^+$ T cells that maintain immune homeostasis. Critically, these cells can interrupt immune surveillance against cancer in healthy individuals, prevent the development of effective antitumor immunity in tumor-bearing patients, and promote tumor progression (Shitara and Nishikawa, 2018). Thus, the differences in tissue infiltration observed for both immune cell subsets in PPV-bearing vs. PPV-free tumors can be explained by PID-induced mechanisms of tumor development.

Of note, inherited genetic mutations have been shown to affect cancer development in those under 40; however, their role in cancer development in those around the age of 50 has not been demonstrated. Nevertheless, identifying the relationship between suspected cancer-predisposing PID-associated mutations is crucial for cancer prevention and the development of enhanced therapeutic approaches. In particular, we note that immunoglobulin replacement therapy is a standard treatment for PID, and it can be used as a prophylactic immune therapy for patients who are expected to have a high incidence of cancer caused by PID (Krivan et al., 2017).

In conclusion, this study presents a comprehensive link between germline mutations in PID-associated genes and cancer, which advances our understanding of both immune surveillance and the mechanisms by which cancer is induced by PID. With this information, it may be possible not only to treat cancer, but also to develop aggressive prophylactic treatments that can replace surgical resection and are unlikely to involve recurrent relapse, which is a fundamental problem of current cancer treatments.

REFERENCES

Abu-Duhier, F.M., Goodeve, A.C., Wilson, G.A., Care, R.S., Peake, I.R., and Reilly, J.T. (2001). Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. British journal of haematology *113*, 983-988.

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics *Chapter 7*, Unit7 20.

Ariga, T., Igarashi, T., Ramesh, N., Parad, R., Cicardi, M., and Davis, A.E., 3rd (1989). Type I C1 inhibitor deficiency with a small messenger RNA resulting from deletion of one exon. J Clin Invest *83*, 1888-1893.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. Behav Brain Res *125*, 279-284.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., *et al.* (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782-795.

Bousfiha, A., Jeddane, L., Picard, C., Ailal, F., Bobby Gaspar, H., Al-Herz, W., Chatila, T., Crow, Y.J., Cunningham-Rundles, C., Etzioni, A., *et al.* (2018). The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. J Clin Immunol *38*, 129-143.

Centeno, I., Blay, P., Santamaria, I., Astudillo, A., Pitiot, A.S., Osorio, F.G., Gonzalez-Arriaga, P., Iglesias, F., Menendez, P., Tardon, A., *et al.* (2011). Germ-line mutations in epidermal growth factor receptor (EGFR) are rare but may contribute to oncogenesis: a novel germ-line mutation in EGFR detected in a patient with lung adenocarcinoma. BMC Cancer *11*, 172.

Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M., and Alizadeh, A.A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol *1711*, 243-259.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., *et al.* (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.

Daugherity, E.K., Balmus, G., Al Saei, A., Moore, E.S., Abi Abdallah, D., Rogers, A.B., Weiss, R.S., and Maurer, K.J. (2012). The DNA damage checkpoint protein ATM promotes hepatocellular apoptosis and fibrosis in a mouse model of non-alcoholic fatty liver disease. Cell Cycle *11*, 1918-1928.

Dohner, H., Estey, E.H., Amadori, S., Appelbaum, F.R., Buchner, T., Burnett, A.K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R.A., *et al.* (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood *115*, 453-474.

Dohner, H., Weisdorf, D.J., and Bloomfield, C.D. (2015). Acute Myeloid Leukemia. The New England journal of medicine *373*, 1136-1152.

Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet *8*, 689-698.

Engels, E.A., Pfeiffer, R.M., Fraumeni, J.F., Jr., Kasiske, B.L., Israni, A.K., Snyder, J.J., Wolfe, R.A., Goodrich, N.P., Bayakly, A.R., Clarke, C.A., *et al.* (2011). Spectrum of cancer risk among US solid organ transplant recipients. JAMA *306*, 1891-1901.

Fathi, A.T., and Chen, Y.B. (2011). Treatment of FLT3-ITD acute myeloid leukemia. American journal of blood research *1*, 175-189.

Filipovich, A.H., Mathur, A., Kamat, D., and Shapiro, R.S. (1992). Primary immunodeficiencies: genetic risk factors for lymphoma. Cancer Res *52*, 5465s-5467s.

Fish, E.N. (2008). The X-files in immunity: sex-based differences predispose immune responses. Nat Rev Immunol *8*, 737-744.

Frohling, S., Schlenk, R.F., Breitruck, J., Benner, A., Kreitmeier, S., Tobis, K., Dohner, H., Dohner, K., and leukemia, A.M.L.S.G.U.A.m. (2002). Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. Blood *100*, 4372-4380.

Gilliland, D.G., and Griffin, J.D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood *100*, 1532-1542.

Grimwade, D., Hills, R.K., Moorman, A.V., Walker, H., Chatters, S., Goldstone, A.H., Wheatley, K., Harrison, C.J., and Burnett, A.K. (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood *116*, 354-365.

Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., de Wind, A., Ravoet, M., Le Buanec, H., Sibille, C., Manfouo-Foutsop, G., *et al.* (2013). CD4(+) follicular helper T cell infiltration predicts breast cancer survival. J Clin Invest *123*, 2873-2892.

Guimaraes, J., Amaral, O., and Sa Miranda, M.C. (2003). Adult-onset neuronopathic form of Gaucher's disease: a case report. Parkinsonism Relat Disord *9*, 261-264.

Hayward, A.R., Levy, J., Facchetti, F., Notarangelo, L., Ochs, H.D., Etzioni, A., Bonnefoy, J.Y., Cosyns, M., and Weinberg, A. (1997). Cholangiopathy and tumors of the pancreas, liver, and biliary tree in boys with X-linked immunodeficiency with hyper-IgM. J Immunol *158*, 977-983.

Hirota, S., Nishida, T., Isozaki, K., Taniguchi, M., Nishikawa, K., Ohashi, A., Takabayashi, A., Obayashi, T., Okuno, T., Kinoshita, K., *et al.* (2002). Familial gastrointestinal stromal tumors associated with dysphagia and novel type germline mutation of KIT gene. Gastroenterology *122*, 1493-1499.

How, J., Sykes, J., Gupta, V., Yee, K.W., Schimmer, A.D., Schuh, A.C., Minden, M.D., Kamel-Reid, S., and Brandwein, J.M. (2012). Influence of FLT3-internal tandem duplication allele burden and white blood cell count on the outcome in patients with intermediate-risk karyotype acute myeloid leukemia. Cancer *118*, 6110-6117.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang, D.S., Tao, H.Q., He, X.J., Long, M., Yu, S., Xia, Y.J., Wei, Z., Xiong, Z., Jones, S., He, Y., *et al.* (2015). Prevalence of deleterious ATM germline mutations in gastric cancer patients. Oncotarget *6*, 40953-40958.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kersey, J.H., Spector, B.D., and Good, R.A. (1974). Cancer in children with primary immunodeficiency diseases. J Pediatr *84*, 263-264.

Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. Nat Rev Cancer *1*, 157-162.

Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A *68*, 820-823.

Krivan, G., Jolles, S., Granados, E.L., Paolantonacci, P., Ouaja, R., Cisse, O.A., and Bernatowska, E. (2017). New insights in the use of immunoglobulins for the management of immune deficiency (PID) patients. Am J Clin Exp Immunol *6*, 76-83.

Lee, S., Wu, M.C., and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. Biostatistics *13*, 762-775.

Leone, P., Vacca, A., Dammacco, F., and Racanelli, V. (2018). Common Variable Immunodeficiency and Gastric Malignancies. Int J Mol Sci *19*.

Levis, M., Allebach, J., Tse, K.F., Zheng, R., Baldwin, B.R., Smith, B.D., Jones-Bolin, S., Ruggeri, B., Dionne, C., and Small, D. (2002). A FLT3-targeted tyrosine kinase inhibitor is cytotoxic to leukemia cells in vitro and in vivo. Blood *99*, 3885-3891.

Levis, M., and Small, D. (2003). FLT3: ITDoes matter in leukemia. Leukemia *17*, 1738-1752.

Li, H. (2011). Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics *27*, 718-719.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Livak, K.J. (1999). Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genetic analysis : biomolecular engineering *14*, 143-149.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lu, C., Xie, M., Wendl, M.C., Wang, J., McLellan, M.D., Leiserson, M.D., Huang, K.L., Wyczalkowski, M.A., Jayasinghe, R., Banerjee, T., *et al.* (2015). Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun *6*, 10086.

Lumsden, J.M., McCarty, T., Petiniot, L.K., Shen, R., Barlow, C., Wynn, T.A., Morse, H.C., 3rd, Gearhart, P.J., Wynshaw-Boris, A., Max, E.E., *et al.* (2004). Immunoglobulin class switch recombination is impaired in Atm-deficient mice. J Exp Med *200*, 1111-1121.

Lyu, X., Yang, J., Wang, X., Hu, J., Liu, B., Zhao, Y., Guo, Z., Liu, B., Fan, R., and Song, Y. (2016). A novel BCR-ABL1 fusion gene identified by next-generation sequencing in chronic myeloid leukemia. Mol Cytogenet *9*, 47.

Machet, L. (1998). [Genetic diseases on the Internet: OMIM]. Ann Dermatol Venereol *125*, 645.

Majhail, N.S., Brazauskas, R., Rizzo, J.D., Sobecks, R.M., Wang, Z., Horowitz, M.M., Bolwell, B., Wingard, J.R., and Socie, G. (2011). Secondary solid cancers after allogeneic hematopoietic cell transplantation using busulfan-cyclophosphamide conditioning. Blood *117*, 316-322.

Mayor, P.C., Eng, K.H., Singel, K.L., Abrams, S.I., Odunsi, K., Moysich, K.B., Fuleihan, R., Garabedian, E., Lugar, P., Ochs, H.D., *et al.* (2018). Cancer in primary immunodeficiency diseases: Cancer incidence in the United States Immune Deficiency Network Registry. J Allergy Clin Immunol *141*, 1028-1035.

McGrath-Morrow, S.A., Gower, W.A., Rothblum-Oviatt, C., Brody, A.S., Langston, C., Fan, L.L., Lefton-Greif, M.A., Crawford, T.O., Troche, M., Sandlund, J.T., *et al.* (2010). Evaluation and management of pulmonary disease in ataxia-telangiectasia. Pediatr Pulmonol *45*, 847-859.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010a). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010b). The Genome Analysis

Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

McKinnon, P.J. (2004). ATM and ataxia telangiectasia. EMBO Rep *5*, 772-776.

Mohammadi, J., Liu, C., Aghamohammadi, A., Bergbreiter, A., Du, L., Lu, J., Rezaei, N., Amirzargar, A.A., Moin, M., Salzer, U., *et al.* (2009). Novel mutations in TACI (TNFRSF13B) causing common variable immunodeficiency. J Clin Immunol *29*, 777-785.

Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T., and Misawa, S. (1996). Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia *10*, 1911-1918.

Nordling, C.O. (1953). A new theory on cancer-inducing mechanism. British journal of cancer *7*, 68-72.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., *et al.* (2008). Genes mirror geography within Europe. Nature *456*, 98-101.

O'Donnell, M.R., Abboud, C.N., Altman, J., Appelbaum, F.R., Coutre, S.E., Damon, L.E., Foran, J.M., Goorha, S., Maness, L.J., Marcucci, G., *et al.* (2011). Acute myeloid leukemia. Journal of the National Comprehensive Cancer Network : JNCCN *9*, 280-317.

Oertelt-Prigione, S. (2012). The influence of sex and gender on the immune response. Autoimmun Rev *11*, A479-485.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., *et al.* (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics *81*, 559-575.

Racanelli, V., and Rehermann, B. (2006). The liver as an immunological organ. Hepatology *43*, S54-62.

Rahman, N. (2014). Realizing the promise of cancer predisposition genes. Nature *505*, 302-308.

Rieux-Laucat, F., and Casanova, J.L. (2014). Immunology. Autoimmunity by haploinsufficiency. Science *345*, 1560-1561.

Rodrigues, F., Davies, E.G., Harrison, P., McLauchlin, J., Karani, J., Portmann, B., Jones, A., Veys, P., Mieli-Vergani, G., and Hadzic, N. (2004). Liver disease in children with primary immunodeficiencies. J Pediatr *145*, 333-339.

Sheikhha, M.H., Awan, A., Tobal, K., and Liu Yin, J.A. (2003). Prognostic significance of FLT3 ITD and D835 mutations in AML patients. The hematology journal : the official journal of the European Haematology Association / EHA *4*, 41-46.

Shin, J., Kim, D., Kim, H.-L., Choi, M., Korbel, J., Yoon, S.-S., and Koh, Y. (2018). Oncogenic effects of germline mutations in lysosomal storage disease genes. bioRxiv, 380121.

Shitara, K., and Nishikawa, H. (2018). Regulatory T cells: a potential target in cancer immunotherapy. Ann N Y Acad Sci *1417*, 104-115.

Southey, M.C., Goldgar, D.E., Winqvist, R., Pylkas, K., Couch, F., Tischkowitz, M., Foulkes, W.D., Dennis, J., Michailidou, K., van Rensburg, E.J., *et al.* (2016). PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet *53*, 800-811.

Tavera-Tapia, A., Perez-Cabornero, L., Macias, J.A., Ceballos, M.I., Roncador, G., de la Hoya, M., Barroso, A., Felipe-Ponce, V., Serrano-Blanch, R., Hinojo, C., *et al.* (2017). Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the ATM gene. Breast Cancer Res Treat *161*, 597-604.

The Genomes Project, C. (2015). A global reference for human genetic variation. Nature *526*, 68-74.

Thompson, D., Duedal, S., Kirner, J., McGuffog, L., Last, J., Reiman, A., Byrd, P., Taylor, M., and Easton, D.F. (2005). Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst *97*, 813-822.

Vajdic, C.M., Mao, L., van Leeuwen, M.T., Kirkpatrick, P., Grulich, A.E., and Riminton, S. (2010). Are antibody deficiency disorders associated with a narrower range of cancers than other forms of immunodeficiency? Blood *116*, 1228-1234.

van Erp, N.P., Eechoute, K., van der Veldt, A.A., Haanen, J.B., Reyners, A.K., Mathijssen, R.H., Boven, E., van der Straaten, T., Baak-Pablo, R.F., Wessels, J.A., *et al.* (2009). Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. J Clin Oncol *27*, 4406-4412.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164.

Yarchoan, R., and Uldrick, T.S. (2018). HIV-Associated Cancers and Related Diseases. N Engl J Med *378*, 2145.

Zhang, J., Nichols, K.E., and Downing, J.R. (2016). Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med *374*, 1391.

국문요약

서론: 이전의 암 관련 생식선 돌연변이 연구는 TP53, APC, BRCA2, NF1, PMS2 및 RB1과 같은 종양 억제 유전자에 집중되어 연구 되어 왔다. 종양 억제 유전자의 경우 병원성 돌연변이의 질병 침투도가 높기 때문에 연구가 용이하여 많은 연구자로 하여금 수행되었다. 이 때문에 암 관련 생식선 연구는 종양 억제 유전자에 초점을 맞추었고 이러한 이유로 종양 억제 유전자 이외의 유전자의 생식선 돌연변이의 역할은 여전히 불분명하다. 그러나, 종양 발달에 영향을 미치는 비 종양 억제 유전자의 존재 가능성은 무시할 수 없다. 생식선 돌연변이로 존재하는 이러한 비 종양 억제 유전자에 의한 종양 형성의 메커니즘은 종양억제 유전자의 종양 발달메커니즘과 다른 방식으로 종양생성에 기여할 것으로 생각된다. 이 연구에서는, 비 종양 억제 유전자의 생식선 돌연변이의 두 가지 다른 유형에 대해 논의 하고자 한다. 첫 번째는 AML 환자의 종양 유전자(Oncogene)의 생식선 돌연변이이고, 두 번째는 선천성 면역 결핍 유전자 돌연변이와 종양 위험 사이의 상관 관계이다.

연구방법: 서울대병원을 내원한 76명의 급성골수성백혈병 환자의 엑솜유전체 데이터가 종양 유발 유전자(Oncogene)의 생식세포 돌연변이 분석에 사용되었다. 2566명의 ICGC 전장 유전체 데이터가 선천성 면역 결핍 유전자와 암 발생의 상관관계를 확인하는데 사용되었으며, 2504명의 1000 게놈 프로젝트 전장 유전체 데이터가 비교 군으로 사용되어 비교 분석 되었다.

결과: 급성골수성백혈병의 발병을 증가시키는 종양 유발 유전자의 생식세포 돌연변이 분석을 위해 FLT3 유전자를 확인 하였다. 76명의 급성골수성백혈

병 환자에서 4개 좌위의 FLT3 생식세포 돌연변이가 발견되었으며 p.D358V 좌위에 해당하는 돌연변이가 급성골수성백혈병 발병을 증가시키는 것이 확인 되었다. 또한 이 돌연변이를 가지고 있는 환자의 경우 좋지 않은 예후를 보였다.

선천성 면역결핍질환의 경우 선천성 면역결핍 질환 유전자는 암 발병을 1.5 배 증가 시켰다. 특히 7개 임종에서 암 위험이 증가하였으며 방광암과 간암의 경우 선천성 면역결핍 질환 유전 변이를 보유한 암환자의 암 발병이 유의하게 일찍 발병하는 것이 확인 되었다.

결론: 우리는 이 논문에서 전에 밝혀지지 않은 두가지 종류의 생식세포 돌연변이와 암 발병위험간의 상관관계를 밝혔다. 첫번째 생식선 돌연변이 유전자는 종양 유전자로 알려져 있는 FLT3돌연변이 이다. 생식선 돌연변이로 존재하는 FLT3의 경우 종양세포의 세포분열을 무한정 촉진 시키는 FLT3 체세포 돌연변이와는 다르게 FLT3기능 저하를 유도하고 대안 세포 신호전달계를 활성 시켜 종양세포의 발달에 기여하는 것으로 판단된다. 두번째 생식선 돌연변이는 선천성 면역 결핍 질환과 관련된 유전자와의 상관관계이다. 선천성 면역결핍과 관련된 유전자의 결핍은 종양 유전자 혹은 종양 억제 유전자들 처럼 직접적인 세포 분열을 촉진 하지 않지만, 면역 감시 효과의 저하로 인해 종양 발생을 촉진 하는 것으로 판단 된다.

이 논문에서 확인된 생식세포 돌연변이뿐만 아니라 밝혀져야 하는 더 많은 알려지지 않은 생식세포 돌연변이가 존재할 것으로 예상된다. 이와 같은 노력은 치료 후 재발하는 암을 극복하기 위한 예방적 치료법을 개발하는데 기여할 것으로 생각된다.

주요어: 생식세포 돌연변이, 유전성 종양, 차세대 염기서열 분석법
학 번: 2016-30613