*Bioinformatics*, 38(8), 2022, 2374–2376
https://doi.org/10.1093/bioinformatics/btac111
Advance Access Publication Date: 18 February 2022
Applications Note

Data and text mining

# PCIG: a web-based application to explore immune–genomics interactions across cancer types

**Anna Pedrola[1,*], Sebastià Franch-Expósito[2], Sara Lahoz[3], Roger Esteban-Fabró[4], Rodrigo Dienstmann[1], Laia Bassaganyas[5,*] and Jordi Camps [3,6,*]**

[1]Oncology Data Science (ODysSey) Group, Vall d'Hebron Institute of Oncology, Barcelona 08035, Spain, [2]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA, [3]Translational Colorectal Cancer Genomics, Gastrointestinal and Pancreatic Oncology Team, IDIBAPS, Hospital Clínic de Barcelona, CIBEREHD, Barcelona 08036, Spain, [4]Liver Cancer Translational Research Group, Liver Unit, IDIBAPS, Hospital Clínic de Barcelona, CIBEREHD, Barcelona 08036, Spain, [5]Department of Medical Genetics, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK and [6]Department of Cell Biology, Physiology and Immunology, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

*To whom correspondence should be addressed.
Associate Editor: Anthony Mathelier

Received on August 12, 2021; revised on December 30, 2021; editorial decision on January 19, 2022; accepted on February 16, 2022

## Abstract

**Motivation:** Genomic alterations can modulate the tumor immunophenotype depending on their nature and tissue of origin. Although this immune–genomic interaction may shape disease progression and response to immunotherapy, the factors governing such dynamics and the influence of each tissue-specific context remain poorly understood.

**Results:** Here, we have developed the PanCancer ImmunoGenomics (PCIG) tool, a web-based resource that provides researchers with the opportunity to mine immunome–genome relationships across several cancer types using data from the Pan-Cancer Analysis of Whole-Genomes (PCAWG) study, which comprises >2,600 samples spanning across 20 different cancer primary sites. PCIG yields an integrative analysis of the crosstalk between somatic genomic alterations and different immune features, thus helping to understand immune response-related processes.

**Availability and implementation:** PCIG is freely available at https://pcig.vhio.net and is supported by all major web browsers. PCIG was developed with Django, which is a Python-based free and open-source framework, and it uses SQL Server as a relational database management system. The code is freely available for download at GitHub https://github.com/AnnaPG/PCIG and in its online supplementary material.

**Contact:** annapedrolagomez@gmail.com or laia.bassaganyas@gmail.com or JCAMPS@clinic.cat

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 Introduction

Cancer genomes play key roles in determining tumor immune features, hence having important implications on disease progression and response to immunotherapy (Thorsson *et al.*, 2018). Somatic genomic alterations may enhance anti-tumor immune activity by enabling the differentiation between self and non-self (tumor) through neoantigen presentation, but they promote immune evasion in later stages of the disease (Litchfield *et al.*, 2021; Mizuno *et al.*, 2021). In fact, the nature of genomic events can affect the genome–immune interaction. Overall, the mutational burden is generally associated with an activated tumor immunome environment and better responses to immunotherapy, whereas high burdens of copy-number alterations (CNAs) often correlate with immune depletion and immunotherapy resistance (Davoli *et al.*, 2017; Tamborero *et al.*, 2018). Moreover, the cancer type and the tissue of origin may also influence the pattern of immune infiltrates (Varn *et al.*, 2017). This complex and dynamic interplay between the cancer genomic landscape and the tumor immune infiltration still remains poorly understood.

To overcome the limited analysis of whole-exome sequencing data, which can hinder the complete view of genomic alterations and complexity, the recently published Pan-Cancer Analysis of Whole Genomes (PCAWG) project (https://dcc.icgc.org/pcawg) includes whole-genome sequencing (WGS) data for more than 2,600 samples from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), spanning up to 20 different cancer types. Importantly, for a subset of 1,300 samples, PCAWG also incorporates whole-transcriptomic data analyzed by RNAseq, providing an exceptional opportunity to comprehensively investigate relationships between genomic alterations and the immune system in primary tumors from different origins (ICGC/

TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020). Nevertheless, WGS data analysis can be costly and time-consuming, limiting its feasibility for rapid and comprehensive exploration.

Here, to examine genome–immunome interactions across a wide spectrum of cancer types, we exploited PCAWG data (https://pcawg.xenahubs.net) to create the PanCancer ImmunoGenomics (PCIG) tool, a web-accessible resource that can be used as a launching platform for clinical and translational research studies to deepen in the exploration of cancer immunogenomics. To this end, we surveyed somatic non-synonymous mutations, CNAs, complex structural variations (SVs), as well as gene expression and clinical classification. Moreover, we created an array of integrative analyses with additional estimated variables, such as the tumor immune composition, the expression of a chemokine-gene signature, the presence of chromothripsis, deletions encompassing the human leukocyte antigen locus, and the levels of broad and focal CNAs. The goal of PCIG is to provide researchers with a fast and easy-to-use tool to visualize the relationships between cancer genomes and immune-related phenotypes to better understand tumor immunogenicity.

## 2 Implementation

PCIG is a web-based interface to explore and visualize the integration of multiple genomic, transcriptomic and immunological features in different cancer types. This tool was created by using the Python-based open-source Django framework. An extended explanation of the PCIG pipeline for data mining and analysis, the genomic and transcriptomic datasets from PCAWG used here, and a description of additional estimated variables are detailed in Supplementary Material. PCIG also provides an extensive User Guide, explaining details for each analysis and different plots presented on the website.

A diagram showing the main parameters and flowchart of the analysis performed by PCIG is depicted in Figure 1A. Briefly, PCIG explores relationships between numerous immune-genomic parameters (Fig. 1A) from 2,658 samples across 40 different cancer types classified based on the primary site (Fig. 1B). Specifically, PCIG employs WGS data to quantify CNA scores and the tumor mutational load per sample, and considers transcriptomic profiles associated with stromal and immune-related genes to perform correlation and integrative analyses to assess the dynamics between these tumor features, and with tumor baseline clinical characteristics (Supplementary Material).

Three main sections are deployed upon selection of primary site and cancer type: Summary, Genomics and Immuno-Genomics. In summary, the main clinical and molecular characteristics are detailed for the selected subset of tumor samples, along with the genomic profiling at the subcytoband level and corresponding parameters analyzed using PCIG's pipeline. Genomics section presents results from the correlative analysis between different genomic variables obtained by WGS data (~2,600 samples), including the number of non-synonymous mutations, broad and focal CNA scores (BCS and FCS, respectively; Franch-Expósito *et al.*, 2020), and the presence of chromothripsis events, indicative of complex SVs (Cortés-Ciriano *et al.*, 2020). Finally, established correlation analyses between genomic variables and different tumor immune metrics computationally derived from transcriptomic data (~1,300 samples) are depicted in the Immuno-Genomics section, including (i) global level of immune and stromal cell infiltrates by ESTIMATE (i.e. ImmuneScore and StromalScore; Yoshihara *et al.*, 2013), (ii) quantification of the four main determinants of tumor immunogenicity by Immunophenoscore (i.e. major histocompatibility complex [MHC]-related antigen processing genes; checkpoints; effector cells; suppressor cells [SC]; Charoentong *et al.*, 2017) and (iii) tumor inflammation assessment through a 12-chemokine gene signature, also associated with the presence of tertiary lymphoid structures, suggestive of a good prognosis in several cancers (Sautès-Fridman *et al.*, 2019).

PCIG provides high comprehensive plots that can be downloaded together with their associated processed data for further analysis. Because of differences in data sources for each cancer type and the limited number of cases in some cohorts, some analyses may require the use of validation datasets.

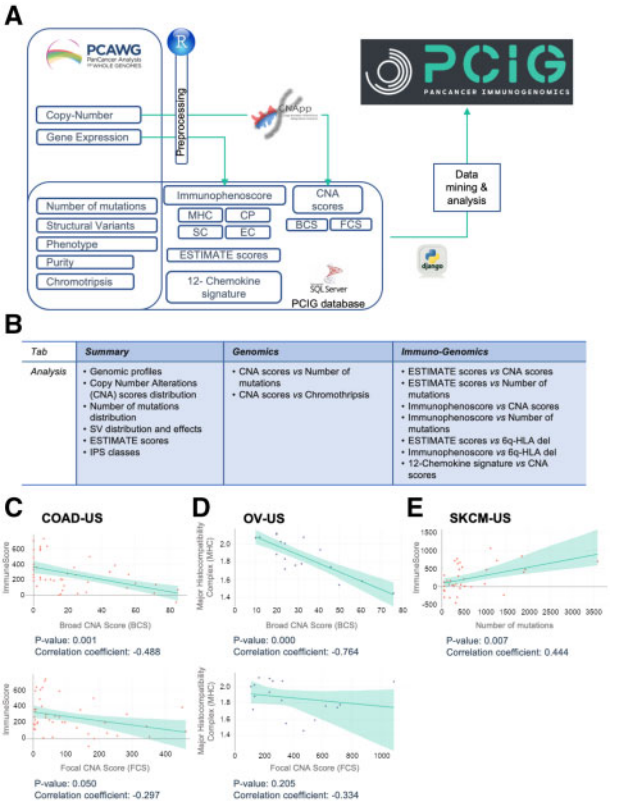


**Fig. 1.** Schematic diagram of PCIG and examples of its performance. (**A**) Flowchart presenting the data sources, types, and analytical tools used by PCIG. (**B**) Detailed summary of the analyses PCIG performs under each tab on the website. (**C**) Correlation plots between ImmuneScore (source: ESTIMATE) and BCS (top) or FCS (bottom) (source: CNApp) using the COAD-US cohort. (**D**) Correlation plot between levels of MHC expression (source: Immunophenoscore) and BCS (top) or FCS (bottom) (source: CNApp) using the OV-US cohort. (**E**) Correlation plot between ImmuneScore (source: ESTIMATE) and the mutational load (source: https://pcawg.xenahubs.net) using the SKCM-US cohort.

## 3 Results and discussion

To exemplify the analytical applicability of PCIG, we explored five datasets: colon adenocarcinoma (COAD-US, $n = 44$), head and neck squamous-cell carcinoma (HNSC-US, $n = 44$), lung adenocarcinoma ($n = 38$), ovarian cancer (OV-US, $n = 42$) and skin cancer (SKCM-US, $n = 37$). Analysis of genomic imbalances showed that ovarian tumors displayed the highest BCS and FCS values (Supplementary Fig. S1A and B), suggesting gross and chromosome-specific aneuploidies. In contrast, the highest values of mutational load were observed in colon cancer, probably due to the presence of POLE-mutated or mismatch repair deficient tumors in the COAD-US dataset (Supplementary Fig. S2). In agreement with previous reports of depleted CD8+ lymphocytic activity in highly aneuploid tumors (Bassaganyas *et al.*, 2020; Davoli *et al.*, 2017), we observed a significant negative correlation between ImmuneScore and BCS or FCS in the majority of cancer types, especially affecting the COAD-US, HNSC-US and OV-US datasets (Fig. 1C and Supplementary Fig. S3A and B). Likewise, tumors with high BCS or FCS such as OV-US exhibited decreased expression of antigen-presenting MHC-related machinery (Fig. 1D and Supplementary Fig. S4), confirming that highly complex genomic tumors bear cold immunophenotypes. Conversely, the presence of a high mutational load observed in skin melanoma and colon adenocarcinoma appeared to be significantly associated with more active immunophenotype profiles (Fig. 1E and Supplementary Fig. S5).

In summary, PCIG correlates cancer genomic traits and immune-related phenotypes, thus helping the interpretation of tumor immunogenicity. In this sense, our analysis of five different cancer

types included in PCAWG further suggests that the tissue of origin and the genomic landscape have an impact on the tumor immune infiltrate. Altogether, PCIG assists in the processing and visualization of large datasets, facilitates exhaustive immune-genomic analyses for hypotheses generation, and displays very complex data in an easy and comprehensible manner.

## Acknowledgements

The authors would like to thank the Vall d'Hebron Institute of Oncology IT for technical support. The study has been developed in part at the Centre Esther Koplowitz from IDIBAPS/CERCA Programme/Generalitat de Catalunya in Barcelona, Spain.

## Funding

This work was supported by the Instituto de Salud Carlos III and co-funded by the European Regional Development Fund (ERDF) [CPII18/00026, PI17/01304], the CIBEREHD and CIBERNED programs from Instituto de Salud Carlos III, the Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya [2017 SGR 1035], and Fundación Científica de la Asociación Española Contra el Cáncer [GCB13131592CAST]. S.L. obtained a PFIS grant from Instituto de Salud Carlos III and co-funded by the European Regional Development Fund (ERDF) [FI18/00221]. R.E.F. is supported by a doctoral training grant from MICINN/MINECO [BES-2017-081286] and a mobility grant from Fundació Universitària Agustí Pedro i Pons. This article is based upon work from COST Action [CA17118] and supported by COST (European Cooperation in Science and Technology).

*Conflict of Interest*: none declared.

## References

Bassaganyas,L. *et al.* (2020) Copy-number alteration burden differentially impacts immune profiles and molecular features of hepatocellular carcinoma. *Clin. Cancer Res.*, **26**, 6350–6361.

Charoentong,P. *et al.* (2017) Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.*, **18**, 248–262.

Cortés-Ciriano,I. *et al.*; PCAWG Consortium. (2020) Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.*, **52**, 331–341.

Davoli,T. *et al.* (2017) Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*, **355**, eaaf8399.

Franch-Expósito,S. *et al.* (2020) CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. *Elife*, **9**, 1–22.

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. (2020) Pan-cancer analysis of whole genomes. *Nature*, **578**, 82–93.

Litchfield,K. *et al.* (2021) Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell*, **184**, 596–614.e14.

Mizuno,S. *et al.* (2021) Immunogenomic pan-cancer landscape reveals immune escape mechanisms and immunoediting histories. Sci. Rep., 11, 15713

Sautès-Fridman,C. *et al.* (2019) Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer*, **19**, 307–325.

Tamborero,D. *et al.* (2018) A Pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin. Cancer Res.*, **24**, 3717–3728.

Thorsson,V. *et al.*; Cancer Genome Atlas Research Network. (2018) The immune landscape of cancer. *Immunity*, **48**, 812–830.e14.

Varn,F.S. *et al.* (2017) Systematic pan-cancer analysis reveals immune cell interactions in the tumor microenvironment. *Cancer Res.*, **77**, 1271–1282.

Yoshihara,K. *et al.* (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.*, **4**, 2612.