

# Editorial on Exome and Genome Retrospective Analysis

Po-Yang Tsou

Department of Medical Informatics, Corpus Christi, Texas, USA

## Editorial

A retrospective investigation is carried out a posteriori based on information about past events. Most of the time, part or all of the information has already been collected and saved in the registry. A retrospective study, unlike prospective studies, does not normally require patients to be followed into the future and often takes less time to complete. Different patient populations can be compared for one or more outcomes in a retrospective study.

In a retrospective study, the desired outcome has already occurred when the investigation begins. Although causal conclusions should not be made, a retrospective study design allows the investigator to establish thoughts about probable links and analyse potential relationships. A retrospective study often uses administrative databases, medical information, or interviews with people who have already been diagnosed with an illness or condition.

Complementary efforts of The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) have recently produced two of the highest quality and most elaborate and reproducible somatic variant call sets from exome and whole genome-level data in cancer genomics, respectively. The motivation for these efforts stems from the notion that "scientific crowd sourcing" and combining mutation callers can provide very strong results.

These two efforts produced variant calls from 10 different callers, namely Radia1, Varscan2, MuSE3, MuTect4, Pindel5,6, Indelocator7, SomaticSniper8 for WES and MuSE, Broad-Pipeline (anchored by MuTect), Sanger-pipeline, German Cancer Research Center pipeline (DKFZ), and SMuFin9, for WGS. Briefly, the PCAWG Consortium aggregated whole genome sequencing data from 2658 cancers across 38 tumor types generated by the ICGC and TCGA projects. These sequencing data were re-analyzed with standardized, high-accuracy pipelines to align to the human genome (reference build hs37d5) and identify germ line variants and somatically acquired mutations.

Of the 885 TCGA samples in ICGC, 746 were included in the latest exome call set produced by both the Multi-Center Mutation Calling in Multiple Cancers (MC3) effort and the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium set. These 746 samples represent a critical benchmark for high-level analysis of similarities and differences between exome and genome somatic variant detection methods.

Imperative to exploiting the molecular vulnerabilities of cancer is the ability to identify potentially actionable genetic alterations. Currently, the most common clinically used sequencing platforms assess a limited number of hotspot mutations in one or greater frequently altered genes (i.e., Cobas testing for BRAF V600E mutations in melanoma). These approaches may

uncover the most extensively validated mutations in several tumor types. However, it is becoming increasingly evident that more extensive analysis of a tumor's genetic landscape is critical in at least the following scenarios. First, driver genes contain activating mutations at non-hotspot locations that confer sensitivity to approved therapies [i.e., BRAF L597 mutations in melanoma] second, crucial alterations that are more prevalent in one malignancy may also predict response to available agents in a distinct tumor type (i.e., BRAF V600E mutation in melanoma and lung cancer). Third, clinically relevant gene fusions are not detected with hotspot testing methods.

Finally, and most importantly, numerous clinical trials involving experimental targeted agents are being conducted. Many agents are now demonstrating signs of efficacy, even in previously recalcitrant gene pathways involving activated RAS, impaired p53, and loss of cyclin-dependent kinase regulation. A significant proportion of patients may therefore be excluded from potentially effective therapeutics based on incomplete genetic profiling.

Cancer is driven by genetic change, and the advent of massively parallel sequencing has enabled systematic documentation of this variation at the whole-genome scale. Here we report the integrative analysis of 2,658 whole-cancer genomes and their matching normal tissues across 38 tumour types from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA). We describe the generation of the PCAWG resource, facilitated by international data sharing using compute clouds.

On average, cancer genomes contained 4–5 driver mutations when combining coding and non-coding genomic elements; however, in around 5% of cases no drivers were identified, suggesting that cancer driver discovery is not yet complete. Chromothripsis, in which many clustered structural variants arise in a single catastrophic event, is frequently an early event in tumour evolution; in acral melanoma, for example, these events precede most somatic point mutations and affect several cancer-associated genes simultaneously. Cancers with abnormal telomere maintenance often originate from tissues with low replicative activity and show several mechanisms of preventing telomere attrition to critical levels. Common and rare germline variants affect patterns of somatic mutation, including point mutations, structural variants and somatic retrotransposition.

A collection of papers from the PCAWG Consortium describes non-coding mutations that drive cancer beyond those in the TERT promoter; identifies new signatures of mutational processes that cause base substitutions, small insertions and deletions and structural variation; analyses timings and patterns of tumour evolution; describes the diverse transcriptional consequences of somatic mutation on splicing, expression levels, fusion genes and promoter activity; and evaluates a range of more-specialized features of cancer genomes.

**How to cite this article:** Po-Yang Tsou, "Editorial on Exome and Genome Retrospective Analysis." *J Health Med Informat* 12 (2021): 385

**\*Address for Correspondence:** Po-Yang Tsou, MD MPH, Department of Medical Informatics, Corpus Christi, Texas, USA

**Copyright:** © 2020 Tsou PY. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Received** 07 September, 2021; **Accepted** 12 September, 2021; **Published** 17 September, 2021