

Whole tumor genomes across cancers

The Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium project, led by the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA), coordinated the sequencing and analysis of 2,583 tumor whole genomes across 38 cancer types. This impressively large project, comprising many working groups focusing on various molecular or genetic features of cancer, has generated valuable data for the cancer research community that will continue to be mined for many years to come.

Cancer genomics, or the genetic elucidation of tumors via genome sequencing, has deepened the knowledge of cancer biology and holds promise for improving therapies. These twin goals of fully understanding cancer processes and harnessing this information to guide clinical treatment, via genomic analysis, require enormous time, energy and resources. Fortunately, the PCAWG Consortium successfully organized an effort to characterize whole-genome sequences from more than 2,500 tumors across 38 cancer types. They have presented unified analysis pipelines and variant calls enabling the analysis of genomic features that are optimally studied through whole-genome sequencing, including non-coding variants and potential drivers, structural variants, retrotransposition events and mitochondrial variants, among others.

The PCAWG [flagship paper](#) published in *Nature* reports the identification of driver mutations (of which 91% of analyzed tumors have at least one) and provides a helpful overview of the main findings from the project. That there are approximately two dozen publications across five journals (*Nature*, *Nature Communications*, *Nature Genetics*, *Nature Biotechnology* and *Communications Biology*) reflects the breadth and depth of the project. There is a convenient [online collection](#) that allows for further exploration of the papers and the data.

Nature Genetics is delighted to feature five of these studies in our pages. Collectively, these studies show the advantage of having

whole-genome sequencing information, because many of these analyses would have been more difficult to comprehensively perform with more targeted (for example, exome) approaches.

The PCAWG Structural Variation Working Group performed [an analysis](#) of the prevalence and characteristics of chromothripsis rearrangements, also known as ‘chromosome shattering’, by using a new approach that they developed called ShatterSeek. They find that chromothripsis patterns are more heterogeneous than previously described, in terms of how many chromosomes and structural variations are involved.

To understand how structural-variation distributions affect chromatin-folding domains in cancer, the same working group as above performed [an analysis](#) of 288,457 somatic structural variations identified across tumors in relation to topologically associated domains (TADs), identifying boundary-altering structural variations. Interestingly, although these rearrangements can have a large effect on the chromatin landscape, their correlation with changes in transcription is low.

Completing the trio of studies published in *Nature Genetics* coming from the PCAWG Structural Variation Working Group is [an analysis](#) of somatically acquired retrotransposition events, which identifies frequent LINE-1 insertions in certain cancer types and shows that these are capable of inducing large deletions and complex rearrangements.

Another particular advantage of analyzing whole-genome sequences is the ability to characterize mitochondrial somatic variants and analyze patterns in relation to cancer type. In [another study](#), the authors identify 7,611 somatic substitutions and 930 indels from mitochondrial DNA in 2,536 cancer samples. They also analyze mutation signatures, finding that mitochondrial-specific and replication-related processes are responsible for most of the mitochondrial somatic mutations.

Finally, the PCAWG Pathogens working group performed [integrated analysis](#) (three different analysis pipelines) of viral associations in tumor sequencing data. They were able to detect viruses in 382 genome and 68 transcriptome sequences, and a high prevalence of known tumor-associated viruses (hepatitis B virus, Epstein–Barr virus and human papilloma virus). They analyze the effects of viral integrations on gene expression, and they observe increased mutations (both single-nucleotide and copy number variants) near viral integration sites.

With its emphasis on openness and sharing, this remarkable resource will continue to pay dividends in the coming years. The PCAWG Consortium should be commended for generating these informative data for the betterment of the cancer research community and, hopefully, for cancer patients themselves. □

Published online: 5 March 2020
<https://doi.org/10.1038/s41588-020-0597-2>