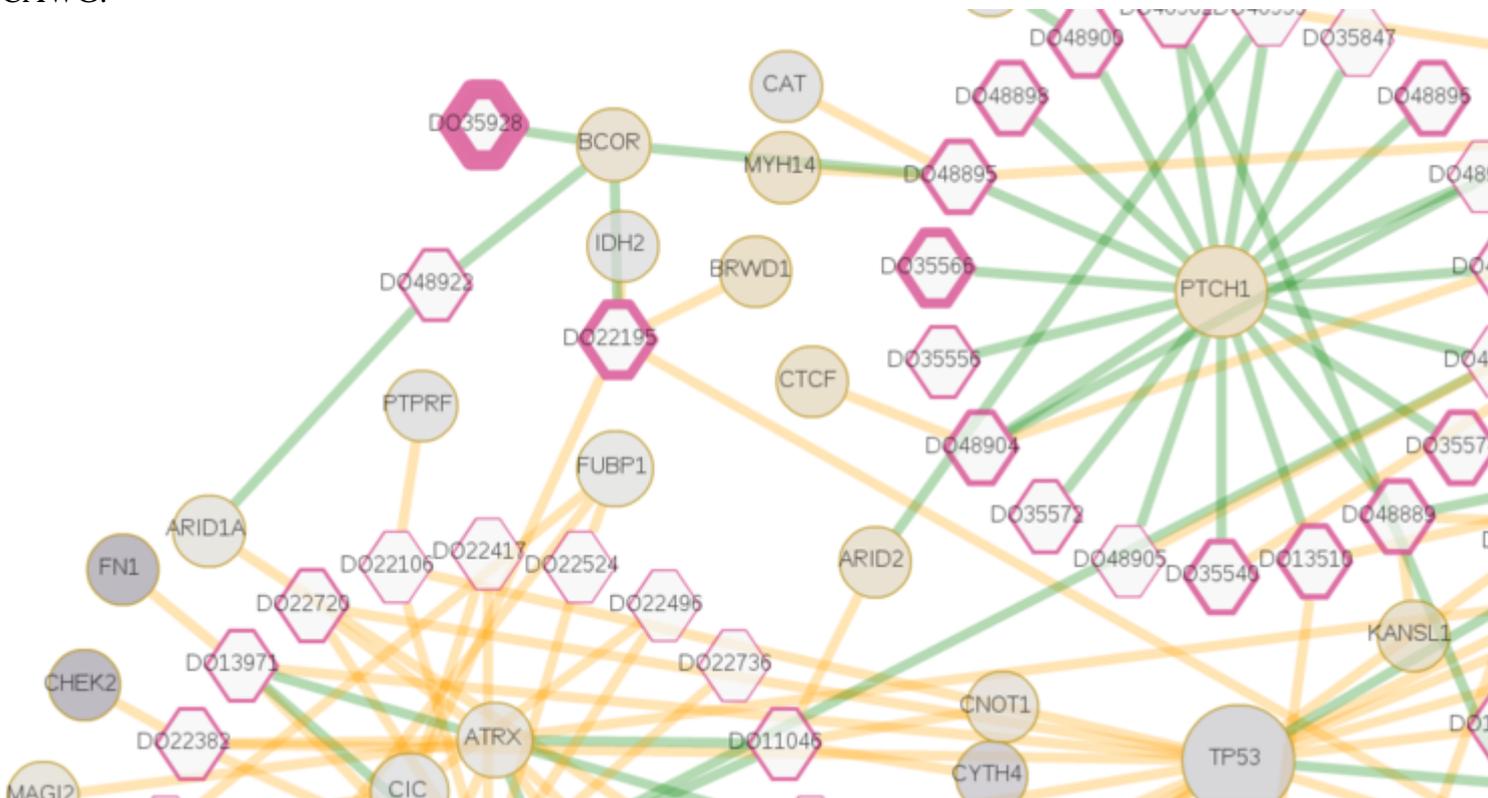


[Inicio](#) > Un grupo de investigadores crean cinco recursos en línea para la exploración, visualización y descubrimiento de datos para el proyecto Pan-Cancer

---

## Un grupo de investigadores crean cinco recursos en línea para la exploración, visualización y descubrimiento de datos para el proyecto Pan-Cancer

El BSC ha liderado la puesta en marcha de la recopilación de herramientas para explorar los datos de PCAWG.



**El centro ha creado una de estas herramientas, PCAWG-Scout, que permite a los usuarios ejecutar sus propios análisis bajo demanda.**

***Nature Communications* publica la investigación**

El Proyecto [Pan-Cancer Analysis of Whole Genomes \(PCAWG\)](#) es un recurso de incalculable valor para comprender la compleja biología del cáncer. Sus 2.658 muestras, elegidas para representar 47 tipos diferentes de tumores, se analizaron en profundidad mediante la secuenciación del genoma completo. El proyecto involucró a más de 1.300 profesionales científicos y médicos de todo el mundo para producir, procesar e interpretar este vasto conjunto de datos, con un tamaño cercano a los 800 TB.

Ahora, cinco instituciones implicadas en el proyecto proporcionan cinco herramientas diferentes de exploración y visualización de datos que permiten a los equipos investigadores examinar este complejo conjunto de datos. En un esfuerzo encabezado por el Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC) y coordinado también por la UCSC (Universidad de California en Santa Cruz), y con la participación de otras instituciones destacadas como el EMBL-EBI (European Bioinformatics Institute), el OICR (Ontario Institute for Cancer Research) y el Harvard Medical School, han producido una guía del usuario de los recursos en línea para la exploración, visualización y descubrimiento de datos para el proyecto Pan-Cancer Analysis of Whole Genomes (PCAWG).

Este documento, publicado en *Nature Communications*, ilustra los diferentes puntos fuertes de sus portales correspondientes, PCAWG-Scout, UCSC Xena, ICGC Data Portal, Expression Atlas y Chromothripsis Explorer, y demuestra cómo se complementan entre sí para comprender más profundamente la biología del cáncer.

El documento detalla casos de uso y análisis para cada herramienta, muestra cómo incorporan recursos externos del ecosistema genómico más grande y demuestra cómo las herramientas se pueden usar juntas para comprender más profundamente la biología de los cánceres.

Todas las herramientas tienen como objetivo agilizar el análisis y la visualización mediante la carga previa de los datos de PCAWG para que los usuarios no necesiten localizar, seleccionar o administrar los datos y hacer que las herramientas sean accesibles a través de una interfaz web. Cada una de estas cinco herramientas también integran otros conjuntos de datos y herramientas de genómica que proporcionan contexto y conocimiento para la interpretación de patrones en los datos de PCAWG, ayudando a este recurso a desarrollar plenamente su potencial.

Miguel Vázquez, líder del grupo Genome informatics del BSC, y responsable principal del artículo, asegura que "los equipos de investigación necesitan herramientas que les muestren estos datos en contexto, para que se revelen patrones interesantes. La naturaleza interactiva de estos portales ofrece la oportunidad de plantear y probar hipótesis. Lo que queríamos destacar en este trabajo es cómo las formas complementarias de sumergirse en los datos que ofrecen estas herramientas revelan fenómenos interesantes que de otro modo estarían ocultos. "

## **Cinco recursos para analizar los datos**

Los cinco recursos que reporta este documento proporcionan una perspectiva y un enfoque diferentes a los datos de PCAWG.

1. [PCAWG-Scout](#), desarrollado por el BSC, permite a los usuarios ejecutar sus propios análisis bajo demanda, incluida la predicción de genes impulsores del cáncer, análisis de expresión génica diferencial, llamadas de variaciones estructurales recurrentes, análisis de supervivencia, enriquecimiento de vías, visualización de mutaciones en una estructura de proteína, detección de firmas mutacionales y recomendaciones para posibles terapias (basadas en el recurso PanDrugs del CNIO).
2. El [ICGC Data Portal](#) sirve como el principal punto de entrada para acceder a todos los datos de PCAWG y también se puede utilizar para explorar mutaciones somáticas simples de consenso de PCAWG, incluidas mutaciones puntuales y pequeños detalles, cada uno por sus frecuencias, patrones de concurrencia, exclusividad mutua y asociaciones funcionales.
3. [UCSC Xena](#) integra diversos tipos de información genómica y fenotípica / clínica a nivel de muestra en la gran cantidad de muestras, lo que permite un examen rápido de los patrones dentro y entre los tipos de datos.

4. El [Chromothripsis Explorer](#) visualiza patrones mutacionales de todo el genoma, con un enfoque en eventos genómicos complejos, por ejemplo, cromotripsis y kataegis. Esto se logra a través de gráficos interactivos de Circos para cada tumor con diferentes pistas que corresponden a variantes de número de copia específicas de alelos, variaciones estructurales somáticas, mutaciones somáticas simples, detalles e información clínica.
5. El [Expression Atlas](#) se centra en los datos de la secuencia de ARN, y respalda las consultas en un contexto de referencia (p. ej., encontrar genes que se expresan en muestras de adenocarcinoma de próstata) o en un contexto diferencial (p. ej., encontrar genes que se expresan de manera insuficiente o excesiva en la próstata adenocarcinomas en comparación con muestras de próstata "normales adyacentes")

## Gestionando 800 TB de datos

La gestión de esta gran cantidad de datos es muy compleja. Moverlos es complicado debido a su volumen y delicado debido a lo sensibles que son en cuanto a la privacidad de los donantes.

Romina Royo, investigadora del grupo INB Computational Node 2 del BSC, afirma: “El análisis que convierte el gran volumen de datos sin procesar en datos procesados es un proceso complejo que requiere enormes recursos computacionales y encuentra numerosas dificultades. Por ejemplo, los tumores que han sufrido eventos catastróficos en la estructura del ADN son un problema para el software de análisis, ya que cuando intentan desenredar los datos se enredan en bucles que pueden disparar en gran medida el tiempo de cálculo”.

Esta primera fase de análisis requirió la colaboración de varios centros de datos; [uno de los cuales fue el BSC](#), que realizó una tarea encomiable siendo uno de los primeros en avanzar resultados, ejecutando gran parte del análisis y almacenando más de 500 TB de datos.

Data Type	# Donors	# Files	Format	Size
SGV	2,715	<a href="#">8,505</a>	VCF	517.27 GB
StGV	2,715	<a href="#">5,668</a>	VCF	7.29 GB
Aligned Reads	2,793	<a href="#">12,169</a>	BAM	<b>794.42 TB</b>
Simple Somatic Mutations	2,715	<a href="#">25,501</a>	VCF	189.99 GB
Copy Number Somatic Mutations	2,715	<a href="#">5,671</a>	VCF	132.62 MB
Structural Somatic Mutations	2,715	<a href="#">14,195</a>	VCF	1.61 GB

Más información sobre las herramientas y sus capacidades, disponible en la página de Visualizaciones y Portales de datos de PCAWG (<http://docs.icgc.org/pcawg>).

**Article:** *A user guide to the online resources for data exploration, visualization, and discovery for the Pan-Cancer Analysis of Whole Genomes project (PCAWG).*

DOI: <https://doi.org/10.1038/s41467-020-16785-6>

**Source URL (retrieved on 14 Jul 2024 - 12:01):** <https://www.bsc.es/es/noticias/noticias-del-bsc/un-grupo-de-investigadores-crean-cinco-recursos-en-l%C3%ADnea-para-la-exploraci%C3%B3n-y-visualizaci%C3%B3n-y>