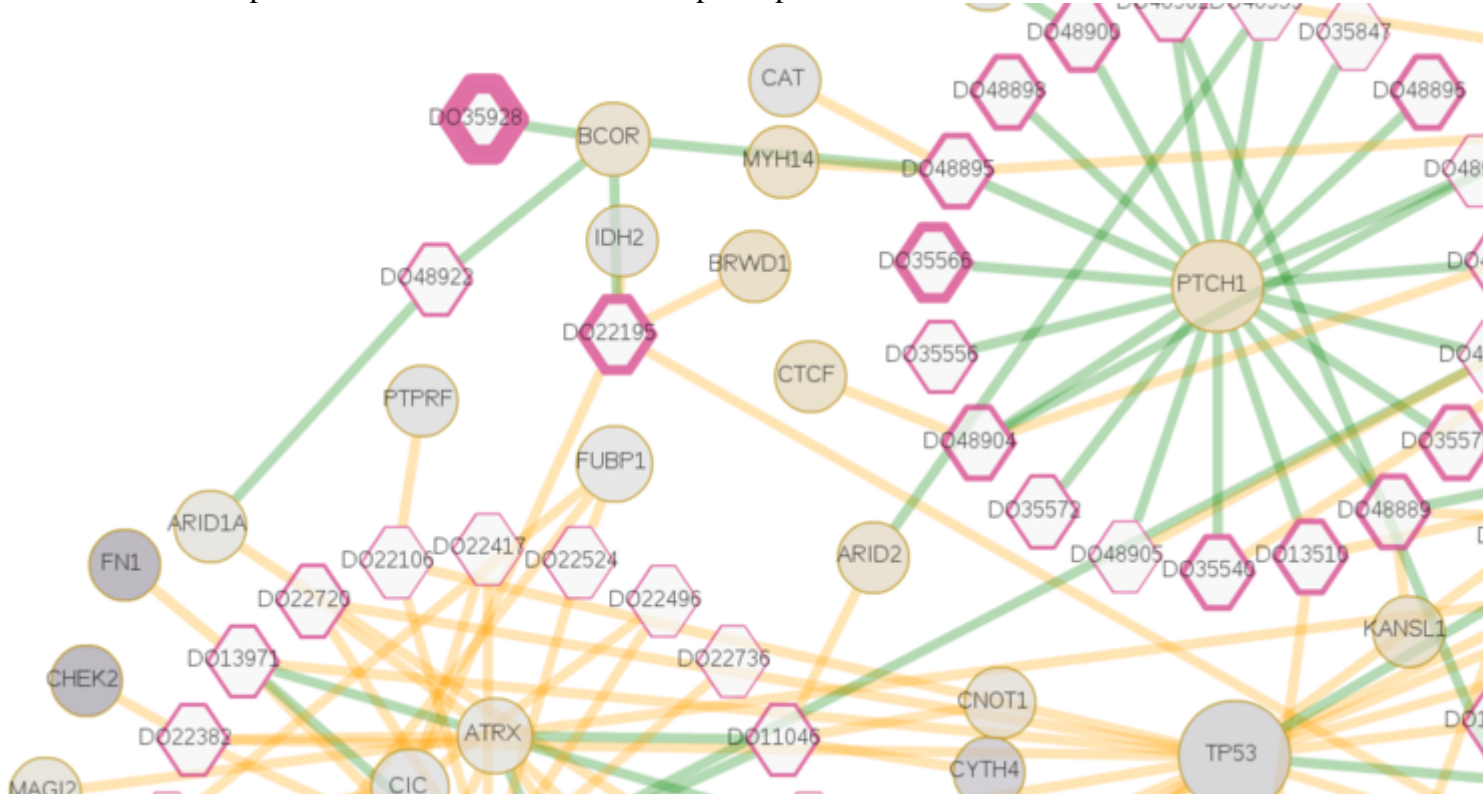


[Inici](#) > Un grup d'investigadors creen cinc recursos en línia per a l'exploració, visualització i descoberta de dades per al projecte Pan-Cancer

## Un grup d'investigadors creen cinc recursos en línia per a l'exploració, visualització i descoberta de dades per al projecte Pan-Cancer

El BSC ha liderat la posada en marxa del recull d'eines per explorar les dades de PCAWG.



**El centre ha creat una d'aquestes eines, PCAWG-Scout, que permet als usuaris executar les seves pròpies anàlisis sota demanda.**

*Nature Communications* publica la investigació

El Projecte [Pan-Cancer Analysis of Whole Genomes \(PCAWG\)](#) és un recurs d'incalculable valor per comprendre la complexa biologia del càncer. Les seves 2.658 mostres, triades per representar 47 tipus diferents de tumors, es van analitzar en profunditat mitjançant la seqüenciació del genoma complet. El projecte va involucrar més de 1.300 professionals científics i mèdics de tot el món per produir, processar i interpretar aquest vast conjunt de dades, prop de 800 TB.

Ara, cinc institucions implicades en el projecte proporcionen cinc eines diferents d'exploració i visualització de dades que permeten als equips investigadors examinar aquest complex conjunt de dades. En un esforç encapçalat pel Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC) i coordinat

també per la UCSC (Universitat de Califòrnia a Santa Cruz), i amb la participació d'altres institucions destacades com l'EMBL-EBI (European Bioinformatics Institute), el OICR (Ontario Institute for Cancer Research) i el Harvard Medical School, han produït una guia de l'usuari dels recursos en línia per a l'exploració, visualització i descoberta de dades per al projecte Pan-Cancer Analysis of Whole Genomes (PCAWG).

Aquest document, publicat a *Nature Communications*, il·lustra els diferents punts forts dels seus portals corresponents, PCAWG-Scout, UCSC Xena, ICGC Data Portal, Expression Atlas i Chromothripsis Explorer, i demostra com es complementen entre si per comprendre més profundament la biologia del càncer.

El document detalla casos d'ús i anàlisi per a cada eina, mostra com incorporen recursos externs de l'ecosistema genòmic més gran i demostra com les eines es poden utilitzar juntes per comprendre més profundament la biologia dels càncers.

Totes les eines tenen com a objectiu agilitzar l'anàlisi i la visualització mitjançant la càrrega prèvia de les dades de PCAWG perquè els usuaris no necessitin localitzar, seleccionar o administrar les dades i fer que les eines siguin accessibles a través d'una interfície web. Cadascuna d'aquestes cinc eines també integren altres conjunts de dades i eines de genòmica que proporcionen context i coneixement per a la interpretació de patrons en les dades de PCAWG, ajudant a aquest recurs a desenvolupar plenament el seu potencial.

Miguel Vázquez, líder del grup Genome informatics del BSC, i responsable principal de l'article, assegura que "els equips d'investigació necessiten eines que els mostrin aquestes dades en context, perquè es revelin patrons interessants. La naturalesa interactiva d'aquests portals ofereix l'oportunitat de plantejar i provar hipòtesis. Allò que volíem destacar en aquest treball és com les formes complementàries de submergir-se en les dades que ofereixen aquestes eines revelen fenòmens interessants que d'una altra manera estarien ocults."

### **Cinc recursos per analitzar les dades**

Els cinc recursos que reporta aquest document proporcionen una perspectiva i un enfocament diferents a les dades de PCAWG.

1. [PCAWG-Scout](#), desenvolupat pel BSC, permet als usuaris executar les seves pròpies anàlisis sota demanda, inclosa la predicció de gens impulsors del càncer, anàlisi d'expressió gènica diferencial, variacions estructurals recurrents, anàlisi de supervivència, enriquiment de vies, visualització de mutacions en una estructura de proteïna, detecció de signatures mutacionals i recomanacions per a possibles teràpies (basades en el recurs PanDrugs de l'CNIO).
2. L'[ICGC Data Portal](#) serveix com el principal punt d'entrada per accedir a totes les dades de PCAWG i també es pot utilitzar per explorar mutacions somàtiques simples de consens de PCAWG, incloses mutacions puntuals i petits detalls, cadascun per les seves freqüències, patrons de concurrència, exclusivitat mútua i associacions funcionals
3. [UCSC Xena](#) integra diversos tipus d'informació genòmica i fenotípica / clínica a nivell de mostra en la gran quantitat de mostres, el que permet un examen ràpid dels patrons dins i entre els tipus de dades.
4. El [Chromothripsis Explorer](#) visualitza patrons mutacionals de tot el genoma, amb un enfocament en esdeveniments genòmics complexos, per exemple, cromotripsis i kataegis. Això s'aconsegueix a través de gràfics interactius de Circos per a cada tumor amb diferents pistes que corresponen a variants de nombre de còpia específiques d'al·lels, variacions estructurals somàtiques, mutacions somàtiques simples, detalls i informació clínica.

## Gestionant 800 TB de dades

La gestió d'aquesta gran quantitat de dades és molt complexa. Moure-les és complicat a causa del seu volum i delicadesa.

Romina Royo, investigadora del grup INB Computational Node 2 del BSC, afirma: "L'anàlisi que converteix el gran volum de dades en estructures comprensibles en l'estructura de l'ADN són un problema per al software d'anàlisi, ja que quan intenten desambolicar les dades, s'acumulen errors".

Aquesta primera fase d'anàlisi va requerir la col·laboració de diversos centres de dades; un dels quals va ser el BSC, que va proporcionar el poder de càlcul i l'espai de dades necessaris.

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

Més informació sobre les eines i les seves capacitats, disponible a la pàgina de Visualitzacions i Portals de dades 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>

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

**Source URL (retrieved on 14 jul 2024 - 16:23):** <https://www.bsc.es/ca/noticies/noticies-del-bsc/un-grup-dinvestigadors-creen-cinc-recursos-en-l%C3%ADnia-lexploraci%C3%B3-visualitzaci%C3%B3-i-descoberta-de-dades>