

## Big Data analysis identifies new cancer risk genes

- Researchers at the [Centre for Genomic Regulation \(CRG\)](#) in Barcelona developed a new method to systematically identify genes contributing to heritable cancer risk
- Their work, which is [published in Nature Communications](#), is a success story for data sharing and openness in science. Just three researchers identified new cancer genes only using publically available data

There are many genetic causes of cancer: while some mutations are inherited from your parents, others are acquired all throughout your life due to external factors or due to mistakes in copying DNA. Large-scale genome sequencing has revolutionised the identification of cancers driven by the latter group of mutations – somatic mutations – but it has not been as effective in the identification of the inherited genetic variants that predispose to cancer. The main source for identifying these inherited mutations is still family studies.

Now, three researchers at the Centre for Genomic Regulation (CRG) in Barcelona, led by the ICREA Research Professor [Ben Lehner](#), have developed a new statistical method to identify cancer predisposition genes from tumour sequencing data. “Our computational method uses an old idea that cancer genes often require ‘two hits’ before they cause cancer. We developed a method that allows us to systematically identify these genes from existing cancer genome datasets” explains Solip Park, first author of the study and Juan de la Cierva postdoctoral researcher at the CRG.

The method allows researchers to find risk variants without a control sample, meaning that they do not need to compare cancer patients to groups of healthy people, “Now we have a powerful tool to detect new cancer predisposition genes and, consequently, to contribute to improving cancer diagnosis and prevention in the future,” adds Park.

The work, [which is published in Nature Communications](#), presents their statistical method ALFRED and identifies 13 candidate cancer predisposition genes, of which 10 are new. “We applied our method to the genome sequences of more than 10,000 cancer patients with 30 different tumour types and identified known and new possible cancer predisposition genes that have the potential to contribute substantially to cancer risk,” says Ben Lehner, principal investigator of the study.

“Our results show that the new cancer predisposition genes may have an important role in many types of cancer. For example, they were associated with 14% of ovarian tumours, 7% of breast tumours and to about 1 in 50 of all cancers. For example, inherited variants in one of the newly-proposed risk genes – NSD1 – may be implicated in at least 3 out of 1,000 cancer patients.” explains [Fran Supek](#), CRG alumnus and currently group leader of the Genome Data Science laboratory at the [Institute for Research in Biomedicine](#) (IRB Barcelona).

### When sharing is key to advance knowledge

The researchers worked with genome data from several cancer studies from around the world, including [The Cancer Genome Atlas \(TCGA\)](#) project and also from several projects having nothing to do with cancer research. “We managed to develop and test a new method

that hopefully will improve our understanding of cancer genomics and will contribute to cancer research, diagnostics and prevention just by using public data,” states Solip Park.

Ben Lehner adds, “Our work highlights how important it is to share genomic data. It is a success story for how being open is far more efficient and has a multiplier effect. We combined data from many different projects and by applying a new computational method were able to identify important cancer genes that were not identified by the original studies. Many patient groups lobby for better sharing of genomic data because it is only by comparing data across hospitals, countries and diseases that we can obtain a deep understanding of many rare and common diseases. Unfortunately, many researchers still do not share their data and this is something we need to actively change as a society”.

### #womeninscience #womeninstem

Postdoctoral researcher Solip Park joined the CRG from South Korea with a Novartis Postdoctoral Fellowship and, more recently, has been awarded a Juan de la Cierva Fellowship and the CRG “[Women Scientists Support \(WOSS\)](#)” grant. This WOSS special grant is an internal initiative of the [CRG Gender Balance Committee](#) to support female scientists who have the ambition and potential to reach a leading position in research.

The CRG is committed to gender equality and, through the Gender Balance Committee aims at eliminating gender bias in recruitment processes, attracting and recruiting female scientists, improving work-life balance, promoting career development and establishing and disseminating gender-sensitive practices.

\*\*\*\*\*

## NOTES TO THE EDITOR

### Reference:

Solip Park, Fran Supek, and Ben Lehner. ‘Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits’ Nature Communications (2018) 9:2601 | DOI: [10.1038/s41467-018-04900-7](https://doi.org/10.1038/s41467-018-04900-7)

**Hi-res images available:** [http://bit.ly/media\\_NatComms\\_ALFRED](http://bit.ly/media_NatComms_ALFRED)



CRG researcher Solip Park developed a method to identify new cancer predisposition genes. Credit: CRG.



Three researchers at the Centre for Genomic Regulation (CRG) identified new cancer risk genes only using public available data. Data sharing is key for genomic research to become more open, responsible and efficient. Credit: Jonathan Bailey, NHGRI.



Genomic data allow researchers to deepen our understanding of the genetic causes of cancer. Credit: 1570578, FreePik.

### Funding information:

This work was supported by a European Research Council (ERC) Consolidator grant (616434), the AXA Research Fund, the Spanish Ministry of Economy and Competitiveness (BFU2011-26206 and ‘Centro de Excelencia Severo Ochoa 2013–2017 SEV-2012-0208), the Agència de Gestió d’Ajuts Universitaris i de Recerca (AGAUR), FP7 project 4DCellFate



(277899), the EMBL-CRG Systems Biology Program and the CERCA Programme of the Generalitat de Catalunya. Fran Supek was also supported by FP7 grants MAESTRA (ICT-2013-612944) and InnoMol (FP7-REGPOT-2012-2013-1-316289). Solip Park. was funded by a Postdoctoral Fellowship from Novartis and by the Juan de la Cierva program (MINECO).

**Genomic data information:**

The results published in this study are in part based upon data generated by The Cancer Genome Atlas project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <https://cancergenome.nih.gov/> . We also acknowledge the 1000 genomes project, the Women's Health Initiative, and UK10K as the sources of primary data.

**For further information and interviews, please, contact:**

Laia Cendrós – Press Officer – Centre for Genomic Regulation (CRG)  
[laia.cendros@crg.eu](mailto:laia.cendros@crg.eu) - Tel. +34 93 316 0237 – Mobile +34 607611798