

**PRESS RELEASE**  
Barcelona, 5th June 2018

## **New tool enables big-scale analysis of single cells**

- ***Researchers at the Centro Nacional de Análisis Genómico of the Centre for Genomic Regulation (CNAG-CRG) developed a novel computational tool named BigScale to analyze millions of single cells simultaneously.***
- ***The study, which is published in the June issue of Genome Research, analyzed 1.3 million cells and unraveled an unprecedented heterogeneity in rare cell populations during mouse brain development.***

New research led by [Holger Heyn](#) at the [Centro Nacional de Análisis Genómico](#) of the Centre for Genomic Regulation (CNAG-CRG), presents a sophisticated computational framework to analyze single-cell gene expression levels, scalable to process millions of individual cells. The work, published in the [current issue of the scientific journal \*Genome Research\*](#), shows for the first time a tool capable of analyzing such large single-cell RNAseq dataset. This dramatically extends the limits of single-cell genome research.

All the cells in a human body share the same genome but each cell has the potential to become specific in a tissue or organ due to gene expression. Scientists worldwide are looking at what differentiates one cell from another. One of the current challenges of genome research is to analyze many individual cells in order to find and to identify those differences. The analysis of individual cell using single-cell RNA sequencing has been crucial to face this challenge and revolutionized our understanding of the complexity of tissues, organs and organisms. Looking at gene expression of one cell at a time, scientists are now able to describe a sample's heterogeneity at unprecedented resolution and without prior knowledge of its composition.

Accordingly, large-scale single-cell projects led to the identification of previously unknown cell types and to the drawing of comprehensive cellular maps of organisms. Within the framework of the Human Cell Atlas project, researchers aim at creating an atlas of all cell types that make up a human body. However, such studies create massive amounts of sequencing data and analyzing large datasets are a major challenge.

A group of scientists of the CNAG-CRG, in collaboration with researchers from the University Pompeu Fabra (UPF) and the Spanish Biomedical Research Consortium on Rare Diseases (CIBERER), has now developed an efficient computational framework that enabled processing, analysis and interpretation of such big-scale single-cell experiments. The group illustrated the power of their strategy by analyzing one of the largest single-cell studies with 1.3 million individual cells of the developing mouse brain.

“BigScale is extremely powerful in identifying cell type specific genes, which greatly helps in the downstream interpretation of experiments” says Holger Heyn, CNAG-CRG team leader and senior author of the study. The novelty of the analytic tool named “BigScale” lies in a numerical model that sensitively determines differences between single cells. Having charted how individual cells differ from each other, they can be grouped together into populations of cells to describe the cellular complexity of a given tissue. As virtually all tissues are composed of different cell types and subtypes, such an analysis can guide an unbiased in-depth characterization without initial hypotheses. Differentially expressed marker genes between subpopulation help the researcher to link cells to prior knowledge about the tissue anatomy or to describe the functions of newly discovered cell types.

In addition, the tool was designed to tackle future challenges of large datasets. “The costs to derive single-cell profiles are decreasing and we are seeing studies of increasing cell numbers” adds Dr. Heyn. In this regard, a module in the BigScale workflow enables the analysis of millions of cells through a directed convolution strategy. Here, single-cell transcriptomes from similar cells are merged into index cells, which greatly reduces the amount of data to be processed.

With the new tool in hand the group analyzed one of the largest single-cell gene expression dataset of 1.3 million cells, a resource publically available from 10x Genomics. “BigScale allowed us to look deep into the developmental processes of the mouse brain and to characterize even rare neuronal cell types” commented Giovanni Iacono, the first author of the work. Specifically, the high number of cells enabled the group to zoom into a small transient cell population called Cajal-Retzius cells and to describe major substructures related to distinct differentiation stages, spatial organization and cellular function. “The BigScale framework provides a powerful solution for virtually any species and is even applicable outside the RNA sequencing context” explains Dr. Heyn and adds “We expect it to contribute to the interpretation of large-scale studies, such as the Human Cell Atlas project”.

\*\*\*\*\*

## **NOTES TO THE EDITOR**

---

**Funding information:** This work received funding and has been supported by the Fundación Ramón Areces, the marathon “Todos somos raros, todos somos únicos”, the EU Marie Curie Fellowships, the Instituto de Salud Carlos III, Spanish Ministry of Economy, Industry and Competitiveness, the Catalan Government, and the European Regional Development Fund (ERDF).

**Reference:** Iacono et al. bigScale: an analytical framework for big-scale single-cell data. Genome Research. 2018. DOI: [10.1101/gr.230771.117](https://doi.org/10.1101/gr.230771.117)

---

## About the Centro Nacional de Análisis Genómico (CNAG-CRG)

The CNAG-CRG is based in Barcelona and it is one of the largest European centers in terms of sequencing capacity. It was created in 2009 with the mission to carry out projects in nucleic acid analysis in collaboration with the national and international research community. It is a non-profit organization funded by the Spanish Ministry of Economy and Competitiveness, and the Catalan Government through the Economy and Knowledge Department and the Health Department. Since 2015 it is part of the Centre for Genomic Regulation (CRG).

The center focuses on sequencing and analysis projects in areas such as cancer genetics, rare disorders, host-pathogen interactions, *de novo* assembly and genome annotation, evolutionary studies and improvement of species of agricultural interest, in collaboration with universities, hospitals, research centers and companies in the sector of biotechnology and pharma.

---

## About the Centre for Genomic Regulation (CRG)

The Centre for Genomic Regulation (CRG) is an international biomedical research institute of excellence, founded in December 2000 whose mission is to discover and advance knowledge for the benefit of society, public health and economic prosperity. The CRG believes that the medicine of the future depends on the groundbreaking science of today. This requires an interdisciplinary scientific team focused on understanding the complexity of life from the genome to the cell to a whole organism and its interaction with the environment, offering an integrated view of genetic diseases. Research at the CRG falls into four main areas: gene regulation, stem cells and cancer; cell and developmental biology; bioinformatics and genomics; and systems biology. The CRG also includes the Centro Nacional de Análisis Genómico (CNAG-CRG) that carries out projects in DNA sequencing and analysis.

---

### 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)

+34933160237 - +34607611798

**Anna Borrell**

Communication Assistant

Centro Nacional de Análisis Genómico (CNAG-CRG)

[anna.borrell@cnag.crg.eu](mailto:anna.borrell@cnag.crg.eu)

+34 934020580