

|                        |            |
|------------------------|------------|
| <b>Date of the CVA</b> | 17/05/2018 |
|------------------------|------------|

## Section A. PERSONAL DATA

|                                    |                    |                     |    |
|------------------------------------|--------------------|---------------------|----|
| Name and Surname                   | Ana Conesa Cegarra |                     |    |
| DNI                                | 22973218J          | Age                 | 49 |
| Researcher's identification number | Researcher ID      | A-9559-2014         |    |
|                                    | Scopus Author ID   |                     |    |
|                                    | ORCID              | 0000-0001-9597-311X |    |

### A.1. Current professional situation

|                       |  |            |  |
|-----------------------|--|------------|--|
| Institution           | Centro de Investigación Príncipe Felipe  |            |  |
| Dpt. / Centre         |  |            |  |
| Address               |  |            |  |
| Phone                 | (0034)<br>963289680<br>- 3012  | Email      | <a href="mailto:aconesa@cipf.es">aconesa@cipf.es</a> |
| Professional category | Senior Group Leader  | Start date | 2007   |
| UNESCO spec. code     | 240401 - Biostatistics   |            |  |
| Keywords              | Multivariate analysis; Computational biology; Genetics transcription; Genetics expresión of regulations; Epigenetics; Genomics; Software; Dataprocessing |            |  |

### A.2. Academic education (Degrees, institutions, dates)

| Bachelor/Master/PhD             | University                         | Year |
|---------------------------------|------------------------------------|------|
| Molecular Microbiology          | University of Leiden               | 2001 |
| Agricultural Engineering Degree | Polytechnic University of Valencia | 1993 |

### A.3. General quality indicators of scientific production

98 Scientific Papers. Google Scholar Citations: 15000

H-index: 42

Conference Talks: 75

Projects as PI: 18

EU/USA projects: 5/4 Total raised ~ 10M euros

PhD thesis: 5 (plus 8 in progress)

Group members in the last 5 years: 35

Visiting students/postdocs last 5 years: 59

## Section B. SUMMARY OF THE CURRICULUM

I am head the Genomics of Gene Expression Lab at the Centro de Investigaciones Príncipe Felipe, Valencia (Spain) and Professor of Bioinformatics at the Microbiology and Cell Science Department at the University of Florida in Gainesville. I graduated as Agricultural Engineer at the Polytechnical University of Valencia in 1993 and obtained my PhD in Molecular Microbiology at the University of Leiden in the Netherlands. After a short appointment as bioinformatics project leader at TNO Quality of Life (The Netherlands). I obtained a Ramon y Cajal award and joined the Valencia Agricultural Research Institute (Spain) in 2003. I moved into CIPF in 2007 and became Senior Group Leader in 2010. In August 2014 I was recruited by the University of Florida through their Pre-eminence hires program and since I enjoy a double appointment CIPF/UF.

I am interested in understanding functional aspects of gene expression at the genome-wide level, across different organisms and in relation to pathological processes. My group has developed statistical methods and software tools that analyze the dynamics aspects

transcriptomes, integrate these with other types of molecular data and annotate them functionally, with a special focus on Next Generation Sequencing (NGS) data. I am creator of popular bioinformatics software such Blast2GO, Paintomics, maSigPro, NOISeq, Qualimap, SQANTI, with tens of thousand users world-wide. I have published nearly 100 research papers that have received more than 14,500 citations and I have an h-index of 42. My labs are well funded through private, national and international funding agencies including H2020, Marie Curie, NIH and USDA. I am scientific coordinator of two large EU projects: the STATegra (11 partners) project on multiomics data integration and the Marie Curie Action DEANN for creating a NGS network with American countries (16 partners), and WP4 leader in ChromE (12 partners). I have (co)-organized numerous bioinformatics and NGS conferences –including the Major Conference in the Computational Biology field ISMB-, and delivered specialized bioinformatics courses in over 10 countries in 5 continents with over 500 attendants. My group is the Valencia Node of the Spanish National Bioinformatics Institute (INB). I am also co-founder of Biobam Bioinformatics S.L., a start-up company that commercializes user-friendly software for biologists. My most current research interests are the development of statistical methods for multiomics data integration for systems medicine, the creation of tools for the analysis of third-generation sequencing data, and the understanding of isoform function.

## Section C. MOST RELEVANT MERITS (ordered by typology)

### C.1. Publications

- 1 Scientific paper.** Nueda, MJ.; et al. 2018. Identification and visualization of differential isoform expression in RNA-seq time series. *Bioinformatics*. 34-3, pp.524-526.
- 2 Scientific paper.** Tardaguila, M.; et al. 2018. SQANTI: extensive characterization of long read transcript sequences for quality control in full-length transcriptome identification and quantification *Genome Research*. 28-3, pp.396-411.
- 3 Scientific paper.** García-Moliner, V; et al. 2018. The SAGA/TREX-2 subunit Sus1 binds widely to transcribed genes and affects mRNA turnover globally *Epigenetics & Chromatin*. 11-13.
- 4 Scientific paper.** Merino, GA.; Conesa, A.; Fernández, EA.2017. A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies. *Briefings in bioinformatics*.
- 5 Scientific paper.** Newman, JRB.; et al. 2017. Disease-specific biases in alternative splicing and tissue-specific dysregulation revealed by multitissue profiling of lymphocyte gene expression in type 1 diabetes. *Genome research*. 27-11, pp.1807-1815.
- 6 Scientific paper.** Ramirez, RN.; et al. 2017. Dynamic gene regulatory networks of human myeloid differentiation *Cell Systems*. Elsevier. 4-4, pp.416-429.
- 7 Scientific paper.** Zhu, Q.; et al. 2017. GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text. *Bioinformatics*.
- 8 Scientific paper.** Conesa, A.; et al. 2016. A Survey of Best Practices for RNA-seq Data Analysis *Genome Biology*. 17-13.
- 9 Scientific paper.** Furio-Tari, P.; et al. 2016. spongeScan: A web for detecting microRNA binding elements in lncRNA sequences *Nucleic Acids Research*. 44-W1, pp.W176-W180.
- 10 Scientific paper.** Tarazona, S.; et al. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package *Nucleic Acid Research*. 43-21, pp.140.
- 11 Scientific paper.** Conesa-Zamora, P.; et al. 2015. Methylome profiling reveals functions and genes which are differentially methylated in serrated compared to conventional colorectal carcinoma *Clinical Epigenetics*. 7-1, pp.101.
- 12 Scientific paper.** Su, ZQ.; et al. 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium *Nature Biotechnology*. 32-9, pp.903-914.
- 13 Scientific paper.** Munro, SA.; et al. 2014. Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures *Nature Communications*. 5, pp.5125.
- 14 Scientific paper.** Garcia-Alcalde, F.; et al. 2012. Qualimap: evaluating next-generation sequencing alignment data *Bioinformatics*. 28-20, pp.2678-2679.

- 15 **Scientific paper.** Tarazona, S.; et al. 2011. Differential expression in RNA-seq: A matter of depth. *Genome Research*. 21-12, pp.2213-2223.

## C.2. Participation in R&D and Innovation projects

- 1 The Next Systems Biology: statistical methods for multiomics systems biology Generalitat Valenciana. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/01/2016-31/12/2019. 190.774 €.
- 2 NOVEL METHODS FOR THE NEW CHALLENGES IN THE ANALYSIS OF HIGH-THROUGHPUT SEQUENCING DATA Ministerio de Economía y Competitividad. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/01/2016-31/12/2018. 220.000 €.
- 3 A Novel Antimicrobial Approach To Combat Huanglongbing USDA-SCRI. Graciela Lorca. (University of Florida). 01/10/2015-30/09/2018. 122.000 €.
- 4 DEVELOPING A EUROPEAN AMERICAN NGS NETWORK (DEANN) MARIE CURIE ACTIONS-PEOPLE. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/01/2014-31/12/2017. 900.000 €.
- 5 Integrative Analysis of TEDDY data to improve T1D diagnosis Helmsley Charitable Trust. Ana Conesa Cegarra. (University of Florida). 01/09/2015-31/08/2017. 600.000 €.
- 6 AYUDA COMPLEMENTARIA para proyecto STATegra personal investigador Consellería de Educación de la Comunitat Valenciana. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/07/2015-31/12/2015. 9.600 €.
- 7 DEVELOPMENT OF COMPUTATIONAL APPROACHES FOR THE CHARACTERIZATION AND FUNCTIONAL ANNOTATION OF LONG-NON-CODING RNA (Annot-lncRNA) Ministerio de Economía y Competitividad. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/01/2013-31/12/2015. 90.000 €.
- 8 USER-DRIVEN DEVELOPMENT OF STATISTICAL METHODS FOR EXPERIMENTAL PLANNING, DATA GATHERING, AND INTEGRATIVE ANALYSIS OF NEXT GENERATION SEQUENCING, PROTEOMICS AND METABOLOMICS DATA Framework Programme 7. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/10/2012-01/10/2015. 6.000.000 €.
- 9 DESARROLLO DE NUEVOS CONCEPTOS Y HERRAMIENTAS BIOINFORMÁTICAS DE NUEVA GENERACIÓN PARA LA PRIORIZACIÓN DE GENES CANDIDATOS EN ENFERMEDADES Y LA ELABORACIÓN DE LAS CORRESPONDIENTES ESTRATEGIAS TERAPÉUTICAS Generalitat Valenciana. Joaquin Dopazo Blazquez. (Centro de Investigación Príncipe Felipe). 01/01/2010-31/12/2014. 244.800 €.
- 10 SECUENCIACIÓN, GENOTIPADO Y DESARROLLO DE HERRAMIENTAS GENÓMICAS PARA LA MEJORA DE LOS CÍTRICOS (CITRUSEQ) Ministerio de Ciencia e Innovación. Investigación. Manuel Talon. (Instituto Valenciano de Investigaciones Agrarias). 29/09/2010-29/09/2014. 355.696 €.
- 11 GENÓMICA Y TRANSCRIPTÓMICA DE LAS RUTAS DE DETOXIFICACIÓN EN DROSOPHILA Ministerio de Ciencia e Innovación. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 15/12/2010-14/12/2013. 53.800 €.
- 12 TÉCNICO DE APOYO PARA REDES TRANSCRIPCIONALES CONTROLADORAS DE VIRULENCIA EN HONGOS FILAMENTOSOS PATÓGENOS Consellería de Educacion, Cultura y Deporte, Generalitat Valenciana. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 15/06/2012-15/11/2013. 9.500 €.
- 13 ANÁLISIS DE GEMELOS DISCORDANTES PARA INVESTIGAR LA CORRELACIÓN ENTRE ALTERACIONES EN EXPRESIÓN Y METILACIÓN DE DNA EN LUPUS ERITEMATOSO SISTÉMICO GentxGent. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/10/2012-30/09/2013. 20.500 €.
- 14 AYUDA COMPLEMENTARIA PARA PATHOGENOMICS - METABOLÓMICA E INTERACTÓMICA DE LA RELACIÓN HUESPED-PATÓGENO Consellería de Educación, Cultura y Deporte, Generalitat Valenciana. (Centro de Investigación Príncipe Felipe). 01/01/2012-31/12/2012. 12.000 €.
- 15 EXPLORING NOVEL GENOME-TRANSCRIPTOME RELATIONSHIPS BY NEXT GENERATION SEQUENCING APPROACHES Ministerio de Economía y Competitividad. (Centro de Investigación Príncipe Felipe). 01/02/2010-01/02/2012. 130.000 €.

- 16 PATHOGENOMICS: REDES TRANSCRIPCIONALES CONTROLADORAS DE LA VIRULENCIA EN HONGOS FILAMENTOSOS PATOGENOS Ministerio de Ciencia e Innovación. Investigación. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/02/2009-01/02/2012. 92.000 €.
- 17 PATHOMICS: METABOLÓMICA E INTERACTÓMICA DE LA RELACIÓN HUESPED-PARÁSITO Ministerio de Ciencia e Innovación. Investigación. (Centro de Investigación Príncipe Felipe). 01/02/2009-01/02/2012. 136.000 €.
- 18 DEVELOPMENT OF TOOLS OF NEW GENERATION FOR GENE EXPRESSION DATA ANALYSIS AND IMPLEMENTATION IN THE IMPROVED GEPAS PLATFORM Fundación Genoma España. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 05/04/2004-05/04/2008. 304.363 €.
- 19 SISTEMAS INTEGRADOS INTELIGENTES EN CÍTRICOS (CITRUS-SII) Conselleria de Empresa, Universidad y Ciencia, Generalitat Valenciana. Ana Conesa Cegarra. (Centro de Investigación Príncipe Felipe). 01/01/2006-31/12/2007. 19.975 €.

### C.3. Participation in R&D and Innovation contracts

- 1 Bioinformatic adviser INMEGEN. PI: Ana Conesa Cegarra. 01/01/2014-01/01/2015. 4.800 €.
- 2 Estudio Bioinformático datos transcriptómica Igenomix. PI: Ana Conesa Cegarra. 01/01/2014-P1Y. 40.000 €.
- 3 Medical Genome Project Junta de Andalucía. PI: Joaquín Dopazo. 01/01/2010-P2Y.
- 4 Blast2GO- obtained functional annotations of the Agilent Gene Expression Microarrays, and the bioinformatics pipeline for further annotations Roslin Insitute. PI: Ana Conesa Cegarra. 01/01/2009-P1Y. 4.000 €.
- 5 Secuencage a grande echelle. Appela porpositon 2003 INRA-Gobierno Frances. PI: Manuel Talon. 01/01/2003-P1Y.

### C.4. Patents

- 1 Software developed: ASCA-genes. Analysis of multifactorial gene expression data (R package)
- 2 Software developed: Blast2GO. Functional annotation & analysis of novel sequence data (Java desktop)
- 3 Software developed: NOISeq. Differential expression analysis of NGS data (R package)
- 4 Software developed: Paintomics. Integration of metabolomics and gene expression data (Web site)
- 5 Software developed: Qualimap. Quality analysis of mapped NGS data (Java desktop)
- 6 Software developed: RGmatch. Linking genomics regions to gene models by NGS data (Python script)
- 7 Software developed: SEA. Analysis of serial gene expression data (Web site)
- 8 Software developed: SQANTI. Structural and Quality Analysis of Transcripts Isoforms (Python script)
- 9 Software developed: STATegraEMS. Experiment Management System for multi-omics experiments
- 10 Software developed: maSigPro. Analysis of time-series gene expression data (R package)
- 11 Software developed: spongeScan. Search for miRNA Multiple Recognition Elements in lncRNAs
- 12 Software developed: tappAS. Functional profiling at the isoform resolution (Java desktop)