**FINAL PUBLISHABLE SUMMARY REPORT**

*1.1 This section should normally not exceed 2 pages.*

The DEANN project is a network formation initiative that involves eight research institutions and universities from five EU countries (United Kingdom, Sweden, Spain, France and Italy) and equivalently, eight research entities from five American countries (Argentina, Mexico, Brazil, Chile and the United States) (Figure 1). The overall goal of this initiative is to strengthen research partnership among project participants by developing a shared scientific knowhow in the field of Next Generation Sequencing (NGS) data analysis. NGS is today the core of the genomics technologies and is predicted that the genomics market will reach 23,300 M € by 2022. While generation of genomics data is currently fast and relatively inexpensive, analysis and knowledge extraction from the data is today the major bottleneck for the efficient application of the genomics paradigm to address disease and well-being questions. Through their exchange programme, the DEANN project strives to reach the following objectives:

**Figure 1. DEANN logo**

- To **reinforce the collaboration** among partner organizations, both between EU and American partners and within EU partners, with the aim of developing long-term research partnerships.
- To nurse the **scientific excellence** of the project participants in the field of NGS data analysis, with the aim of improving their position and competitiveness at the international level.
- To increase the competence of the project partners in applied genomics research with the aim of reaching **translational opportunities** in established and emerging economies.
- To **improve education**, innovation and international orientation of PhD candidates and post-docs.

The DEANN initiative achieves these objectives by creating a network of carefully chosen bioinformatics research labs and putting them to work together in an exchange, training and transference program that specifically serves the goals of the project. The consortium members collectively represent a diversity of expertise domains (Biomedicine, biotechnology and bioinformatics), strategic regional positioning, strong international presence, comprehensive training capability and long-term projection. The DEANN exchange program counts with WPs in the most trendy and translational topics of Next Generation Sequencing and purposely includes movement of management staff and intensive training modules both on scientific and translational skills.

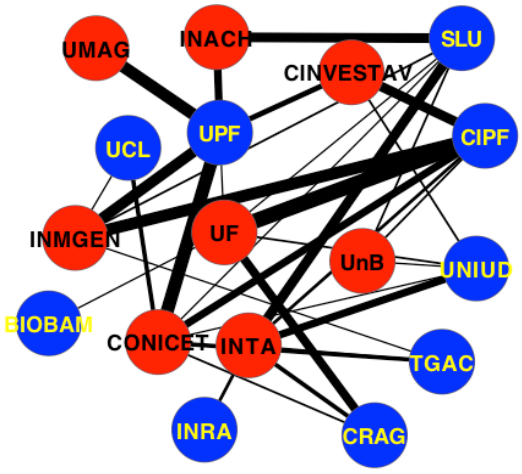


**Figure 2 DEANN scientific network. Nodes are reserach center (blue from Europe and read from America). The thickness of the edge is proportional to the number of exchange visits.**

The DEANN project started in 2014, and by 2018, a total of 141 visits between Europe and America took place, for a total of more than 270 months and involving about 100 scientists. This has created a strong research and collaborative network between partners of the project (Figure

2). As a direct result of the DEANN exchange program 18 scientific papers were published while another 11 are under review or drafting.

Since the beginning of the project, the DEANN network has worked in training and research within different front-end topics of NGS and computational biology research. Scientific activities have been developed around existing research topics of the involved laboratories, who, through the exchange opportunities have been able to expand the scope and quality of their research goals.

Globally, the DEANN network has addressed NGS analysis challenges on the analysis and annotation of genomes, this includes assembly, structural and single variant annotation, annotation of functional elements and identification of haplotypes. They have also investigated dynamic and functional genomics aspects, including gene expression, epigenetic modifications, microRNAs, cis-regulatory elements, etc. DEANN researchers have worked on the integration of different types of omics data to understand complex biological systems, and have worked on metagenomics projects that study the interaction of microorganisms in environmental and human samples. Finally, the DEANN network has developed general-purpose algorithms and software tools for the statistical analysis of NGS data and the acceleration of computational jobs.

We highlight some of the relevant topics of the DEANN network research activities.

An important part of the research effort has been devoted to species of agricultural and environmental relevance. Hence researchers from Argentina, Italy and UK together studied in detail the genome and expression of the sunflower and Eucalyptus. They assembled new genomes, identified structural variants, associated them to important agricultural traits and created pan-genomes, which are the share genetic components of different varieties of the same species. Another important collaboration effort between Argentinean and Spanish scientists was in the study of the genome of several Drosophila species adapted to different toxic environments. Novel methods for genome assembly combining multiple sequencing technologies were developed and extensive transcriptional studies in these species identified genes and pathways that confer environmental adaptations and serve as models for understanding toxicity responses in higher organisms. Researchers from Chile were most interested in ocean species, in particular the Antartic sea urchin, and in collaboration with Spanish, UK and Sweden scientists established genome structure and the role of microRNAs in the fish biology. Mexican and Spanish scientist studied parasitic nematode *Steinernema carpocapsae, and* proposed novel co-evolutionary theories.

Regarding human research, an important part of the effort was dedicated to study genome composition and variation between Mexican and European populations and how this diversity is relevant to disease causative genes. They study the effect of Single Amino Acid Substitutions in the whole genome in terms of stability and impact in protein function to identify targets for drug development. Mexican and Spanish researchers study the genome variation of the parasite *Leshmania mexicana* and how this is relevant for disease treatment.

Finally, DEANN researchers developed a number of user-friendly software tools for the Quality Control and analysis of functional genomics data, such as Qualimap, Paintomics and SQANTI. These tools are able to analyze a wide variety of different types of NGS technologies, apply quality control steps, discover new proteins and integrate them to create complex omics models. Through collaboration between CONICET and UPF, DEANN investigators created *VarQ: a tool for the structural analysis of Human Protein Variations*. Work between Italy and USA researchers lead to new statistical methods for the analysis of allele specific expression while

the collaboration between Spain and Mexico resulted in methodologies for the functional analysis of microRNAs. The tools are available to the scientific community as free software.

The activities and results of the DEANN program are relevant to a wide set of stakeholders, including genomics scientists, clinicians, ecologists, agricultural researchers and bioinformaticians

More info at website: http://bioinfo.cipf.es/deann/