Programme’s research areas and strategic goals

The aim of the Structural Biology Programme (SBP) is to provide mechanistic understanding at the molecular level of how proteins and macromolecular complexes related to cancer function. The ultimate goal is to use the new mechanistic insights, and the solved structures, to help guide the search for new compounds and molecules that could interfere with the function of these complexes. Our current research focuses on the study of protein kinases and the DNA damage response. SBP undertakes this thanks to the multiple technologies available via CNIO’s Units and Research Groups and through the constant lookout for synergies with other CNIO groups. This year a focus was placed on setting up cryo-electron microscopy (cryo-EM) methods at SBP, a revolutionary technology to observe individual macromolecules at high resolution that is reshaping biological research.

Summary of milestones & major achievements during 2018

It has been approximately a year now that 3 new Groups started at SBP, many positive changes have taken place since then. Looking back, we can be proud of our many achievements in 2018. The Groups and Units at SBP have made a collective effort to promote collaborations between them, as well as with other groups at the CNIO. The new Groups have set up their labs and new equipment has been acquired. In many cases these resources have been shared between groups, thereby fostering a spirit of scientific interactions, as well as rationalising the use of our resources.

We have made every effort to make our new Programme known around the world. Several top researchers in the field of Structural Biology came to the CNIO as speakers, including John L. Rubinstein (University of Toronto, Canada), and Kiyoshi Nagai (LMB-MRC, Cambridge, UK), among others. In addition, we actively advertised the Programme abroad; we visited the LMB-MRC (Cambridge, UK), the Institute of Cancer Research (ICR, London, UK) and the Max Planck Institute of Biochemistry (Munich, Germany). We are also organising a CNIO Frontiers Meeting to be held in the spring of 2019 for which we will be bringing in some of the most renowned experts in Structural Biology of the DNA damage response.

On June 25th, SBP was granted a competitive project to access the Electron Bio-Imaging Centre (eBIC) at the Diamond Light Source (UK). Our project, titled “Stop cancer - structural studies of macromolecular complexes involved in cancer by cryo-EM”, will allow access to this top cryo-EM facility. The Electron Microscopy Unit forms part of this project; the intention is to allow the access to eBIC for CNIO groups that are outside the SBP through a collaboration with the Electron Microscopy Unit. As part of our achievements, the 2 new Junior Groups were awarded ‘Ramon y Cajal’ research contracts and also both obtained grants from the Spanish Government, an excellent boost to start off their independent research. In 2018, SBP published the first atomic structure obtained using cryo-electron microscopy methods. In addition, Daniel Lietha published new insights into the mechanisms involved in the activation of Focal Adhesion Kinase.

Finally, I would like to acknowledge the work of Daniel Lietha, Junior Group Leader at SBP. Over the last few years, Daniel has significantly contributed to the structural and mechanistic understanding of Focal Adhesion Kinase, and other signalling pathways important for proliferation, adhesion and cancer survival. Daniel has also invested considerable effort to look for new ligands to inhibit Focal Adhesion Kinase using structural approaches. On top of all this, Daniel has been an exceptional colleague, always available to help everyone. Daniel Lietha is now moving to CSIC as a ‘distinguished researcher’ and we all wish him and his team the best in this new phase of their scientific career.

In summary, during 2018, SBP has set up the infrastructure, the collaborations and the working environment that we will need in the coming years to push our research, improve our connectivity with CNIO, and to contribute to provide a mechanistic and structural understanding of processes relevant in cancer.

“SBP endeavours to define structures of macromolecules relevant in cancer in order to provide mechanistic understanding, which is a first step towards new therapies. In 2018, SBP determined its first cryo-EM structures.”
Activation and assembly of many protein complexes implicated in cancer, such as kinases and polymerases, require the assistance of HSP90, a molecular chaperone. Thus, HSP90 inhibitors are being evaluated as anticancer agents.

HSP90 is needed for the activation and stability of the PI3-kinase-like kinases (PIKKs), including mTOR, ATM and ATR that regulate the DNA damage response and cell growth. Surprisingly, these kinases require the action of HSP90 but working in concert with the R2TP/Prefoldin-like (R2TP/PFDL) complex. R2TP/PFDL is the most complex HSP90 co-chaperone yet described. R2TP/PFDL contains multiple subunits and growing evidence links this complex to cancer.

Yet, how all these processes work is largely unknown. We are using cryo-electron microscopy (cryo-EM) to fully understand the molecular mechanisms of R2TP/PFDL and to bring us a step closer to designing strategies to interfere with PIKK assembly and activation.

“How kinase complexes implicated in cancer are assembled by HSP90 and R2TP is unclear. The structure of R2TP brings us a step closer to mechanistic understanding and the design of anticancer strategies.”
**RESEARCH HIGHLIGHTS**

**Cryo-EM and structure of macromolecular complexes in cancer**

A defining feature of our Group is our interest in understanding the structural and molecular mechanisms of macromolecular complexes involved in the DNA damage response. For this, we use mostly biochemical and molecular biology tools in combination with cryo-electron microscopy (cryo-EM). Cryo-EM is used to visualise large macromolecular complexes, to observe their flexibility and motions, and to build atomic models. Cryo-EM is especially helpful for complex and flexible assemblies, which are typically difficult to crystallise. The structural and functional information provides mechanistic details to help understand the DNA damage response, and it is an input for the design of new strategies to interfere with these processes.

The Group is currently working on several complexes implicated in the response to DNA damage, but this year our main area of focus was the characterisation and understanding of how HSP90 and the R2TP co-chaperone function to assemble large macromolecular complexes of relevance in cancer.

How cells build protein interactions in protein kinase complexes

Assembly, activation and cellular stability of a growing list of macromolecular complexes, many of which are relevant in cancer, require the assistance of molecular chaperones. Among these, the kinases of the PI3-kinase-like family (PIKKs) function as part of large multi-subunit complexes that require HSP90 for assembly. The PIKK family comprises proteins such as ATM, ATR and DNA-PKcs, implicated in DNA repair and DNA damage signalling, and mTOR, which controls cell growth. These kinases interact with other proteins in order to function properly and be active, as in the mTOR complex 1 (mTORC1) or ATR-ATRIP. Building these protein interactions needs sufficient conformational freedom to interact with a diversity of clients.

Together, our findings provide the first structural view of human R2TP, an essential component of the HSP90-mediated assembly of mTORC1, ATR-ATRIP and other complexes of the PIKK family. Our structures also highlight important differences between the human complex and the much simpler homologs found in yeast. We have discovered an intricate architecture of the human R2TP complex, providing a flexible tether for HSP90, needed to cope with the assembly of multiple and diverse macromolecular complexes. A structural view of how HSP90 and its co-chaperone assists the assembly of proteins involved in cancer will bring us a step closer to the potential design of new anticancer strategies.

**How novel structures inform understanding of complement function.**


**Figure 1** Selected views of the human R2TP complex as observed by cryo-EM. (a) Several domains can be localised, and the flexibility of the HSP90-binding regions in RAPAP3 is detected.

**Figure 2** Cryo-EM map of the human R2TP complex, showing the C-terminal domain (blue colour) bound to the C-terminal domain of RPAP3 (yellow colour), as seen from the top (a) and side (b). (c) Detail of the ADP-binding site. The quality of the cryo-EM density, represented in mesh, is sufficient to detect the ADP and side chains of residues in the binding site for nucleotides.
Our Group studies regulatory mechanisms of key signalling switches controlling growth and adhesion signals, which regulate important cellular processes such as cell proliferation, migration and survival. We use structural techniques, such as X-ray crystallography and electron microscopy, in combination with biochemical and functional studies to understand these mechanisms at atomic detail and to rationalise how oncogenic events result in their deregulation. The structural understanding allows us to design potential anti-cancer therapeutics that interfere with oncogenic deregulation.

We focus on growth and adhesion signalling systems that interact and are regulated by specific lipids in the plasma membrane. Specifically, we pursue 2 main questions:

1. How are adhesion signals in focal adhesion complexes triggered by membrane interactions?
2. How are the levels of specific lipids regulated by the SHIP lipid phosphatase to control growth signals?

“Using structure-based design, we generated the first irreversible and sub-nanomolar inhibitor targeting adhesion signals that trigger cancer invasion.”

\[ \text{OVERTVIEW} \]

**RESEARCH HIGHLIGHTS**

Focal Adhesion Kinase (FAK) is a key regulator of adhesion signals and localises into a signalling layer on the plasma membrane in focal adhesion complexes. We previously discovered that FAK interacts with PIP2 lipids in focal adhesions and this triggers its activation by inducing FAK oligomerisation, conformational changes that facilitate its autophosphorylation, Src recruitment and FAK phosphorylation by Src. Currently, we are studying the atomic architecture of FAK oligomers bound to PIP2 membranes by electron microscopy (EM). We have obtained a 5.9 Å map, which reveals the mode of oligomerisation and large membrane induced rearrangements of FAK’s regulatory FERM and kinase domains (FIGURE). The observed conformation suggests that FAK adopts a ‘preactivated’ primed state when bound to the membrane. We are further investigating how force, induced at focal adhesion sites by actomyosin contraction, can induce changes to these structures to fully activate focal adhesion signalling. We utilise these mechanistic insights to discover highly specific allosteric FAK inhibitors. We employ a fragment based approach to identify allosteric ligands and then use structure based drug design to develop these fragments into inhibitory lead compounds.

SHIP phosphatases remove the 5-phosphate from PIP3 and thereby, like PTEN, negatively regulate PIP3 levels in the plasma membrane. Despite their importance, little is known about mechanisms of SHIP regulation. We previously solved a crystal structure containing the catalytic and C2 domains of SHIP2, which, together with extensive biochemistry and cell biology experiments, showed how the C2 domain induces catalytic activation of SHIP2. Currently, we are studying the role of the PH domain flanking the catalytic domain. We find that the domain binds the PIP2 substrate and PIP2 product, and that this binding allosterically further activates SHIP. Together, this shows how the C2 and PH domains concertedly act to recruit SHIP to PIP2 rich membranes in order to adopt a highly active state.

**PUBLICATIONS**


**ANNUAL REPORT 2018**
**OVERVIEW**

Rational and precise targeting of oncogene driven signalling is a crucial and yet outstanding challenge in cancer research today. Understanding the structural and molecular bases of oncogene activation and signalling is key for the design and development of better therapeutics. Our research focuses on the structural and molecular understanding of protein kinase function: how protein kinases are activated and regulated by post-translational modifications and allosteric inputs, and how they assemble into macromolecular protein complexes to transmit signals inside the cell. We put a special emphasis on how these mechanisms are corrupted in cancer and disease due to oncogenic mutations and other oncogenic insults. Crucially, such atomic and molecular information can be translated into the design and development of more potent and specific protein kinase inhibitors, eventually leading to more effective drugs for the treatment of cancer patients.

“Understanding protein kinase function and inhibition for better cancer therapeutics.”

During 2018, we have set up the different experimental systems and techniques needed for the adequate functioning of the lab and have established 3 main research lines:

- Structural and molecular determinants of RET catalytic activity and signalling, both in cis by intrinsic elements and in trans by effector kinases and adaptor proteins.
- Structure-function studies of RET oncogenic variants, i.e. point mutations targeting the kinase domain and oncogenic fusions generated by DNA-rearrangements.
- Structure-based drug-discovery of (allosteric) RET inhibitors.

Furthermore, upon invitation by the journal *Endocrine-Related Cancer*, we contributed to a special issue to commemorate the 25th anniversary of the discovery of the RET proto-oncogene as the cause of Multiple Endocrine Neoplasia type 2 (see publication list).
Mismatch repair

DNA mismatch repair (MMR) is critical for genome stability. The DNA mismatch repair machinery loads onto newly synthesised DNA and searches for mismatches. The recognition of an error in DNA by the MutS protein leads to an ATP-dependent conformational change that transfers MutS into a sliding clamp state. Only this MutS state can activate the MutL ATPase that in turn promotes the cleavage of the DNA for repair. These protein complexes are incredibly dynamic and flexible, and many steps of the cycle have remained elusive to structural analysis. Using cryo-EM, we have captured multiple functional steps and we have studied the conformational changes that these proteins undergo in order to recognise the mismatch and license downstream events that lead to repair. These studies are carried out in collaboration with Titia Sixma (Netherlands Cancer Institute) and Meindert Lamers (Leiden University).

DNA replication & repair - focus on mitochondria

Eukaryotic cells have two genomes: nuclear and mitochondrial. However, how the integrity of the mitochondrial genome is maintained through the equilibrium between DNA replication, repair and degradation, and organelle dynamics, remains unclear. We are interested in understanding these pathways because of their implications for ageing and disease, and in particular, their relation to cancer.

Cryo-electron microscopy (cryo-EM)

Combined with many other approaches already established at the CNIO, we use cryo-EM to study diverse macromolecular complexes involved in cancer. Significant recent technological developments in microscopes, detectors and image processing tools have significantly improved the resolution of the technique, enabling the structural analysis of many elusive macromolecules to an unprecedented level of detail. Last year, we worked together with the Óscar Llorca Group and the EM Unit to bring the cryo-EM facility at the CNIO to a state-of-the-art level. Moreover, we have been awarded access to high-end microscopes at the Biological Electron Bio-Imaging Centre (eBIC) in Oxford (UK). We can now efficiently prepare samples and solve their structures, using the in-house facilities, to a high level of detail.

OVERVIEW

Safeguarding the genetic information is essential to all forms of life. Two key cellular processes keep it free from errors: DNA replication and repair. Importantly, when these do not work correctly, genetic information may be damaged or lost, ultimately leading to disease. Deregulation and malfunction of the protein machinery that safeguards our genome are a hallmark of cancer, but it remains unclear how this happens at the molecular level. The devil is in the detail, and we aim to understand to the highest level of detail what and when things can go wrong with these molecular machines, so we can act on it to correct it and prevent it from happening.

These macromolecules are like real life machines, with intricate mechanisms that enable them to perform their activities. To understand how they work, we use cryo-electron microscopy and biochemistry in an integrative approach. Beyond fundamental research, this structural information provides the necessary detail for drug development.

“...The high-end cryo-electron microscopy setup at the CNIO allows us to look at every detail of the cell’s protein machinery, so that we can understand how it works and intervene.”

Awards and Recognition

- Awarded with a Ramón y Cajal Fellowship, Spanish Ministry of Science, Innovation and Universities.
SPECTROSCOPY AND NUCLEAR MAGNETIC RESONANCE UNIT

Ramon Campos-Olivas
Head of Unit

Technician
Clara M. Santiveri (TS)
Titulado Superior (Advanced Degree)

RESEARCH HIGHLIGHTS

Our Core Unit incorporates a broad range of instrumentation for the biophysical characterisation of biomolecules and their interactions, including spectrophotometers, a fluorimeter, isothermal titration and differential scanning calorimeters, a circular dichrograph, dynamic and multi-angle static light scattering devices, and a surface plasmon resonance (SPR) instrument. Research groups mostly from, but not limited to, the Structural Biology Programme extensively used these technologies throughout 2018. For example, in collaboration with the Experimental Therapeutics - ETP - Programme, we conducted quantitative binding measurements using NMR (see FIGURE) to establish that a cell-active small molecule compound interacts weakly with telomeric double stranded DNA. Thus, telomeric DNA binding appears not to be a significant mode of action of the compound to explain its cellular activity.

The Unit hosts a 700 MHz NMR spectrometer that is well equipped with probes, and a sample changer for running up to 120 samples automatically. This provides the required throughput for the screening of small molecule protein binders (together with the CNIO’s Structural Biology and ETP Programmes), as well as for metabolomics measurements that, this year, were performed in collaboration with the CNIO-Lilly Cell Signalling Therapies Section (from the ETP), the Cell Division and Cancer Group (-CDC- Group, from the Molecular Oncology Programme), as well as the Growth Factors, Nutrients and Cancer and the Epithelial Carcinogenesis Groups (from the Cancer Cell Biology Programme). For example, in collaboration with the CDC Group, we conducted cell media and intracellular metabolite measurements to characterise the metabolic changes associated to the silencing of the Mastl gene.

Collectively, with these and other groups, we implemented sample preparation protocols and developed spectroscopic and analytical tools to characterise the metabolites present in different biological samples.

“In 2018, we identified and quantified interactions of small molecule compounds with tumour-relevant proteins and DNA, thereby contributing to the discovery of possible macromolecular inhibitors, as well as to the understanding of the molecular bases of the cell activity of those compounds.”

OVERVIEW

The Unit unifies the technical and scientific management of Nuclear Magnetic Resonance (NMR) Spectroscopy and other molecular biophysics instrumentation available through the Structural Biology Programme. It provides CNIO researchers with equipment and technical support for a variety of techniques used in biophysical studies of molecules involved in cancer. This includes the application of NMR to the in vitro characterisation of the structure and dynamics of biomolecules (proteins in particular) and their interactions with other biopolymers, as well as with small molecules that could represent initial hits in the drug discovery process or research compounds for biophysical and functional studies. Furthermore, we use NMR to characterise the metabolic profiles of biofluids, cell growth media and cell and tissue extracts from both animal models of cancer and human samples.

“... we conducted cell media and intracellular metabolite measurements to characterise the metabolic changes associated to the silencing of the Mastl gene. Collectively, with these and other groups, we implemented sample preparation protocols and developed spectroscopic and analytical tools to characterise the metabolites present in different biological samples."

Figure NMR study of the interaction of a small molecule-compound with a DNA duplex containing 7 telomeric repeats. (A) Superposition of the aromatic region of the 1H NMR spectrum of the compound (100 µM) recorded after addition of increasing amounts of dsDNA (from bottom to top: 0-8.9 µM). The tilted arrows mark signals changing position upon dsDNA addition, identifying the chemical moieties directly involved in dsDNA binding. In contrast, the 3 signals from the phenyl ring of the compound only experience broadening as a result of complex formation, but not change of position, as the phenyl moiety does not directly contact the dsDNA. (B) Linear variations in the chemical shift (spectral positions) of three signals as a function of added dsDNA indicate weak binding (𝐾_𝐽 > 10 µM).
Bioinformatics is a key discipline for furthering our understanding of the cancer genome and for the future of cancer therapeutics. Bioinformatics-based approaches have the ability to transform the huge amount of biological data into comprehensive models that provide an in-depth understanding of cancer disease and the complex relationships among genotype and phenotype that are needed to identify cancer driver molecular alterations and new therapeutic targets.

The CNIO Bioinformatics Unit (BU) has several goals: i) to provide bioinformatics support with data analysis and interpretation using computational and statistical methods; ii) to achieve genome analysis in cancer patients’ data in order to identify new biomarkers and mechanisms of drug response; iii) to maintain the CNIO’s scientific computing facilities and expertise for the systematic analysis and interpretation of cancer genomes. Importantly, the Bioinformatics Unit is extensively involved in teaching activities – with an important focus on the translational bioinformatics area – to train bioinformatics users and developers. We co-organise the Master in Bioinformatics Applied to Medicine Personalized and Social (BICEM-ENS) as well as advanced Bioinformatics courses for sequencing analysis (visit our web page for a full list of activities).

**SELECTED PUBLICATIONS**

- Tress L, Rodríguez JM, Vázquez J, Tress L. (2018) Vulcanspot workflow: 1) identification of genome-wide vulnerable Gene Dependencies (GD) integrating functional genomics databases; 2) to propose drugs to target GDs following a dual strategy; and 3) therapeutic prioritization as a final output targeting GDs detected in the user’s gene list.
**Research Highlights**

Over the last decade, cryo-EM has emerged as a key technique for studying how biomolecules function and interact. Our 120-kV Spirit G12 EM, equipped with the TVIPS CMOS detector, has cryo-capabilities that enable sample screening and low resolution analysis of standard biological specimens. For high resolution cryo-EM data collection, CNIO’s Structural Biology Programme has been granted access to high-end cryo-EM microscopes at the Electron Bio-Imaging Centre (eBIC) (Oxford, UK) through peer-reviewed Block Allocation Group (BAG) access.

Throughout 2018, the Unit has performed EM experiments with all the research groups from the Structural Biology Programme, as well as with several groups from other CNIO Programmes and outside our Centre. For example, in collaboration with CNIO’s Microenvironment and Metastasis Group, we have contributed to the characterisation of circulating extracellular vesicles from the lymph and plasma of melanoma patients. Our data supported their analysis of lymph-circulating extracellular vesicles for the detection of residual disease and its reappearance in melanoma. The most frequent mutation in amyotrophic lateral sclerosis and frontotemporal dementia patients involves mutation of the C9ORF72 gene, resulting, to a certain extent, in the expression of toxic dipeptide arginine repeats (PR). We have evaluated, by electron microscopy, in collaboration with the CNIO Genomic Instability Group, that the presence of (PR)20 did not affect the in vitro assembly of purified 40S and 60S subunits into 80S particles in the absence of mRNA. Moreover, in collaboration with Dr Iván Ventoso – from the Centro de Biología Molecular ‘Severo Ochoa’ (CSIC-UAM) and the Departamento de Biología Molecular of the Universidad Autónoma de Madrid (UAM) – we used electron microscopy to localise gold-labelled eukaryotic initiation factor-4A (eIF4A) in the ribosomal translation initiation complex 48S. Our results have contributed towards the proposal of a topological model of the scanning ribosomal 43S pre-initiation complex.

---

**Overview**

The Electron Microscopy (EM) Unit is a central core facility as well as a research laboratory. It is available to CNIO researchers and the wider research community, providing investigators with instruments and support for Transmission Electron Microscopy analysis. The Unit offers negative staining and cryo-EM specimen preparation techniques for proteins, protein complexes and vesicles. We also give training to regular users on the use of our equipment and provide further guidance regarding specimen preparation.

“Over the past year, the Electron Microscopy Unit has endeavoured to adapt its facility to better meet the needs of the new members of our Programme, particularly in relation to the cryo-EM technique.”

---

**Publications**

OVERVIEW

The Crystallography and Protein Engineering Unit (XTPEUnit) is a central core facility as well as a research laboratory whose main goal is to supply the requests of the research groups within the CNIO, as well as external groups, by providing on-demand services at different levels. The Unit covers all the services that range from the coding groups, by providing on-demand services at different levels.

The Unit works closely with the Experimental Therapeutics Programme on several projects that have since led to the production of recombinant proteins (full-length human MASTL, full-length mouse TRF1 and human TRF1 dimerization domain) for biochemical experiments. We also focus on structural characterisation by X-ray crystallography in support of drug discovery projects, as in the case of the kinase HASPIN - crystallised in the presence of compounds developed at the Medicinal Chemistry Section (FIGURE) - and other projects like MASTL, DDR1 and the dimerization domain of TRF1 (FIGURE), the latter in strict collaboration with the Telomeres and Telomerase Group. We have continued our work on the production of proteins for the generation of antibodies by the Monoclonal Antibody Unit (Biotechnology Programme), including several cancer-involved proteins such as HJH1, TRT2, NGFR, RCT2, CDK16, PotA, MIDKINE, GPM1, Ly9, CD85A, and HASPIN, among others. The Unit is also engaged in numerous internal collaborations with other CNIO Research Groups and Units, noteworthy are the following: Telomeres and Telomerase, Experimental Oncology, Genomic Instability, Cell Division and Cancer, Molecular Imaging Unit, Microenvironment and Metastasis, Gastrointestinal Cancer Clinical Research Unit, Melanoma, Epithelial Carcinogenesis, Familial Cancer Unit, H120-CNIO Lung Cancer Clinical Research Unit, and the H120-CNIO Haematological Malignancies Clinical Research Unit. Additionally, our Unit has ongoing collaborations with external groups such as the Environmental Biology Department (CNB-CSIC), the Biomedical Application of Radioisotopes Unit (CNB-CSIC), the CNB-CSIC Radiobiology Unit, the Roswell Park Cancer Institute (USA), the Protein Tools Unit (CNB-CSIC), the Biomedical Application of Radioisotopes Unit (CIB-CSIC), the Department of Molecular Engineering (Aarhus University, Denmark), have shown their high tumour inhibition potency. Finally, the Unit is taking part in a project, in collaboration with the CNIO’s Molecular Imaging Unit, to develop new antibody-based positron emission tomography (immunoPET) imaging tools for tumour visualisation.

**“We aim to show how protein information at atomic resolution is crucial for understanding the processes occurring in the biology of cancer.”**

**RESEARCH HIGHLIGHTS**

The Unit works closely with the Experimental Therapeutics Programme on several projects that have since led to the production of recombinant proteins (full-length human MASTL, full-length mouse TRF1 and human TRF1 dimerization domain) for biochemical experiments. We also focus on structural characterisation by X-ray crystallography in support of drug discovery projects, as in the case of the kinase HASPIN - crystallised in the presence of compounds developed at the Medicinal Chemistry Section (FIGURE) - and other projects like MASTL, DDR1 and the dimerization domain of TRF1 (FIGURE), the latter in strict collaboration with the Telomeres and Telomerase Group. We have continued our work on the production of proteins for the generation of antibodies by the Monoclonal Antibody Unit (Biotechnology Programme), including several cancer-involved proteins such as HJH1, TRT2, NGFR, RCT2, CDK16, PotA, MIDKINE, GPM1, Ly9, CD85A, and HASPIN, among others. The Unit is also engaged in numerous internal collaborations with other CNIO Research Groups and Units, noteworthy are the following: Telomeres and Telomerase, Experimental Oncology, Genomic Instability, Cell Division and Cancer, Molecular Imaging Unit, Microenvironment and Metastasis, Gastrointestinal Cancer Clinical Research Unit, Melanoma, Epithelial Carcinogenesis, Familial Cancer Unit, H120-CNIO Lung Cancer Clinical Research Unit, and the H120-CNIO Haematological Malignancies Clinical Research Unit. Additionally, our Unit has ongoing collaborations with external groups such as the Environmental Biology Department (CNB-CSIC), the Biomedical Application of Radioisotopes Unit (CNB-CSIC), the CNB-CSIC Radiobiology Unit, the Roswell Park Cancer Institute (USA), the Protein Tools Unit (CNB-CSIC), the Biomedical Application of Radioisotopes Unit (CIB-CSIC), the Department of Molecular Engineering (Aarhus University, Denmark), have shown their high tumour inhibition potency. Finally, the Unit is taking part in a project, in collaboration with the CNIO’s Molecular Imaging Unit, to develop new antibody-based positron emission tomography (immunoPET) imaging tools for tumour visualisation.

**AWARDS AND RECOGNITION**

- Member of the Board of Directors, Asociación de Universidades de Dirección de España.

**PUBLICATIONS**

- Mortuza GB et al. (2018). CRISPR-based positron emission tomography (PET) imaging biomarker for pancreas cancer therapy. Fit of the ab initio SAXS structures (envelopes in grey) and the corresponding generated trimerbodies models. (chains in blue, magenta and cyan).
Biomedical cancer research is a particularly data-heavy discipline, where key information sources are not only limited to genomic information or raw experimental data. Especially unstructured data, such as the scientific literature, clinical texts, medicinal chemistry patents or patient generated content, constitute a valuable resource for a range of scenarios like drug discovery, interpretation of large scale experimental results, drug repurposing or evidence based medicine. Medical big data approaches are only able to efficiently exploit running texts through the use of natural language processing (NLP) techniques relying on deep learning and artificial intelligence strategies. Our Unit is financed through the Plan for the Advancement of Language Technologies; the aim is to generate resources that can improve the exploitation of biomedical data by means of implementing and evaluating the underlying quality of systems for automatic recognition of medical concepts, generation of specialised neural machine translation models for the medical domain and the implementation of a medical language technology platform and software components for processing Spanish EHRs.

The Biological Text Mining Unit has provided consultancy, guidance and technical support for clinical text mining use cases posed by several healthcare institutions (Hospital Virgen del Rocio, Hospital XII de Octubre, Hospital Son Espases, Hospital Clinic), national and regional health-related agencies (Spanish Medical Agency, Instituto Aragonés de Ciencias de la Salud, Servicio Andaluz de Salud, Fundación TIC Salud Social), and natural language as well as medical informatics academic research groups. The Unit has contributed to benchmarking efforts of clinical text mining systems by organising shared tasks in the context of community challenges organised by the Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN-IberEval) and releasing high quality evaluation datasets. The Unit has published a collection of clinical NLP resources, all freely available at https://zenodo.org/communities/medicalnlp and https://github.com/PlanTL.

In addition to annotation guidelines and Gold Standard corpora for developing and evaluating the quality of systems for automatically detecting biomedical and clinical concepts, the Unit has implemented software tools for automatic medical term recognition and normalisation (CUTEXT), an electronic health record sectionizer, a medical sentence boundary recognition system, a medical text tokenizer, lemmatizer and PoS-tagger. Moreover, we have also contributed to the first Protected Health Information (PHI) masker for the Spanish language, a system for medical negation detection, clinical temporal expression detection based on HeidelTime, a medical machine translation system and word embeddings. These key constituents are being integrated into the clinical NLP pipeline developed by the Unit.

“Language technologies, together with artificial intelligence, are driving the technological transformation of biomedical and clinical data into actionable information at all levels of cancer research.”

**RESEARCH HIGHLIGHTS**

**Data collection**
Collect anonymized discharge reports through a multi-provider health system

**Recommendations**
Based on the previous evaluation, we generate good practice recommendations for medical personnel.

**Variable definition**
The variables to be extracted are those outlined by clinical experts, starting from a smaller set of key variables, being evaluated and incorporated with each iteration.

**Evaluation**
The evaluation of the quality of the process for the discharge reports are key objectives. The project will provide a corpus and associated Gold annotated, releasing them in the context of a community challenge.

**SpaCEx System**

**Figure** Clinical NLP framework for processing electronic health records in Spanish and Catalan.