



RESEARCH CENTER

FIELD

**Digital Health, Biology and Earth**

Activity Report 2018

# Section Application Domains

Edition: 2019-03-07



## COMPUTATIONAL BIOLOGY

1. ABS Project-Team (section vide)	5
2. BEAGLE Project-Team	6
3. BIGS Project-Team	7
4. BONSAI Project-Team (section vide)	8
5. CAPSID Project-Team	9
6. DYLISS Project-Team	11
7. ERABLE Project-Team	12
8. GENSCALE Project-Team	13
9. IBIS Project-Team (section vide)	14
10. LIFEWARE Project-Team	15
11. MORPHEME Project-Team (section vide)	17
12. MOSAIC Team (section vide)	18
13. PLEIADE Team	19
14. SERPICO Project-Team	23

## COMPUTATIONAL NEUROSCIENCE AND MEDICINE

15. ARAMIS Project-Team	25
16. ATHENA Project-Team	26
17. BIOVISION Project-Team	27
18. CAMIN Team (section vide)	28
19. EPIONE Project-Team (section vide)	29
20. GALEN-POST Team	30
21. MATHNEURO Team (section vide)	33
22. MIMESIS Team	34
23. MNEMOSYNE Project-Team	35
24. NEUROSYS Project-Team	36
25. PARIETAL Project-Team	37
26. VISAGES Project-Team	39

## EARTH, ENVIRONMENTAL AND ENERGY SCIENCES

27. AIRSEA Project-Team	40
28. ANGE Project-Team	41
29. CASTOR Project-Team	44
30. COFFEE Project-Team	45
31. FLUMINANCE Project-Team (section vide)	47
32. LEMON Team	48
33. MAGIQUE-3D Project-Team	49
34. SERENA Project-Team	51
35. STEEP Project-Team (section vide)	52
36. TONUS Team	53

## MODELING AND CONTROL FOR LIFE SCIENCES

37. BIOCORE Project-Team	54
--------------------------	----

38. CARMEN Project-Team .....	56
39. DRACULA Project-Team (section vide) .....	57
40. M3DISIM Project-Team .....	58
41. MAMBA Project-Team .....	59
42. MONC Project-Team .....	66
43. NUMED Project-Team (section vide) .....	67
44. REO Project-Team .....	68
45. SISTM Project-Team .....	70
46. XPOP Project-Team .....	71

**ABS Project-Team (section vide)**

## **BEAGLE Project-Team**

# **4. Application Domains**

## **4.1. Domain 1**

Applications concern Functional and Evolutionary Biology, plant, animal and human health. They are not detailed here because the project itself is oriented by its applications, so the description of applications is described along the project in the previous sections.

## BIGS Project-Team

### 4. Application Domains

#### 4.1. Tumor growth-oncology

On this topic, we want to propose branching processes to model appearance of mutations in tumor through new collaborations with clinicians. The observed process is the "circulating DNA" (ctDNA). The final purpose is to use ctDNA as a early biomarker of the resistance to an immunotherapy treatment. It is the aim of the ITMO project. Another topic is the identification of dynamic network of expression. We continue our work on low-grade gliomas. The ongoing collaboration with Montpellier CHU, and a new one with Montreal CRHUM should provide us more data. We initiate as well interactions with researchers from Montreal LIO to extend the previous work. We still have much work to do in modeling to reach our goal of a decision-aid tool for personalised medicine. In the same context, there is a question of clustering analysis of a brain cartography obtained by sensorial simulations during awake surgery.

#### 4.2. Genomic data and micro-organisms population

Despite of his 'G' in the name of BIGS, Genetics is not central in the applications of the team. However, we want to contribute to a better understanding of the correlations between genes trough their expression data and of the genetic bases of drug response and disease. We have contributed to methods detecting proteomics and transcriptomics variables linked with the outcome of a treatment.

#### 4.3. Epidemiology and e-health

We have many works to do in our ongoing projects in the context of personalized medicine with "CHU Nancy". They deal with biomarkers research, prognostic value of quantitative variables and events, scoring, and adverse events. We also want to develop our expertise in rupture detection in a project with APHP for the detection of adverse events, earlier than the clinical signs and symptoms. The clinical relevance of predictive analytics is obvious for high-risk patients such as those with solid organ transplantation or severe chronic respiratory disease for instance. The main challenge is the rupture detection in multivariate and heterogeneous signals (for instance daily measures of electrocardiogram (during 30 minutes), body temperature, spirometry parameters, sleep duration, etc. Other collaborations with clinicians concern foetopathology and we want to use our work on conditional distribution function to explain fetal and child growth. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy).

#### 4.4. Dynamics of telomeres

The telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Trough a beginning collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently data on the distribution of the length of telomeres from blood cells. With some members of Inria team TOSCA, we want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc.); and (3) use these properties to develop new statistical methods. A slot of postdoc position is already planned in the Lorraine Université d'Excellence, LUE project GEENAGE (managed by CHU Nancy).

**BONSAI Project-Team (section vide)**



## CAPSID Project-Team

### 4. Application Domains

#### 4.1. Biomedical Knowledge Discovery

**Participants:** Marie-Dominique Devignes [contact person], Malika Smail-Tabbone, Sabeur Aridhi, David Ritchie, Gabin Personeni, Seyed Ziaeddin Alborzi, Kévin Dalleau, Bishnu Sarker, Emmanuel Bresso, Claire Lacomblez.

This projects in this domain are carried out in collaboration with the Orpailleur Team.

Huge and ever increasing amounts of biomedical data (“Big Data”) are bringing new challenges and novel opportunities for knowledge discovery in biomedicine. We are actively collaborating with biologists and clinicians to design and implement approaches for selecting, integrating, and mining biomedical data in various areas. In particular, we are focusing on leveraging bio-ontologies at all steps of this process (the main thesis topic of Gabin Personeni, co-supervised by Marie-Dominique Devignes and Adrien Coulet from the Orpailleur team). One specific application concerns exploiting Linked Open Data (LOD) to characterise the genes responsible for intellectual deficiency. This work is in collaboration with Pr. P. Jonveaux of the Laboratoire de Génétique Humaine at CHRU Nancy [71], [72]. This involves using inductive logic programming as a machine learning method and at least three different ontologies (Gene Ontology, Human Phenotype Ontology, and Disease Ontology). This approach has also been applied using pattern structure mining (an extension of formal concept analysis) of drug and disease ontologies to discover frequently associated adverse drug events in patients [70]. This work was performed in collaboration with the Centre for BioMedical Informatics Research (BMIR) at Stanford University.

Recently, a new application for biomedical knowledge discovery has emerged from the ANR “FIGHT-HF” (fight heart failure) project, which is in collaboration with several INSERM teams at CHRU Nancy. In this case, the molecular mechanisms that underly HF at the cellular and tissue levels will be considered against a background of all available data and ontologies, and represented in a single integrated complex network. A network platform is under construction with the help of a young start-up company called Edgeleap. Together with this company, we are developing query and analysis facilities to help biologists and clinicians to identify relevant biomarkers for patient phenotyping [48]. Docking of small molecules on candidate receptors, as well as protein-protein docking will also be used to clarify a certain number of relations in the complex HF network.

#### 4.2. Prokaryotic Type IV Secretion Systems

**Participants:** Marie-Dominique Devignes [contact person], Bernard Maigret, Isaure Chauvot de Beauchêne, David Ritchie, Philippe Noël, Aichata Niang.

Prokaryotic type IV secretion systems constitute a fascinating example of a family of nanomachines capable of translocating DNA and protein molecules through the cell membrane from one cell to another [37]. The complete system involves at least 12 proteins. The structure of the core channel involving three of these proteins has recently been determined by cryo-EM experiments [56], [80]. However, the detailed nature of the interactions between the remaining components and those of the core channel remains to be resolved. Therefore, these secretion systems represent another family of complex biological systems (scales 2 and 3) that call for integrated modeling approaches to fully understand their machinery.

In the frame of the Lorraine Université d'Excellence (LUE-FEDER) "CITRAM" project MD Devignes is pursuing her collaboration with Nathalie Leblond of the Genome Dynamics and Microbial Adaptation (DynAMic) laboratory (UMR 1128, Université de Lorraine, INRA) on the discovery of new integrative conjugative elements (ICEs) and integrative mobilisable elements (IMEs) in prokaryotic genomes. These elements use Type IV secretion systems for transferring DNA horizontally from one cell to another. We have discovered more than 200 new ICEs/IMEs by systematic exploration of 72 *Streptococcus* genome. As these elements encode all or a subset of the components of the Type IV secretion system, they constitute a valuable source of sequence data and constraints for modeling these systems in 3D. Another interesting aspect of this particular system is that unlike other secretion systems, the Type IV secretion systems are not restricted to a particular group of bacteria [46].

### 4.3. Protein-RNA Interactions

**Participants:** Isaure Chauvot de Beauchêne [contact person], Antoine Moniot, Bernard Maignet, Maria Elisa Ruiz Echartea, David Ritchie, Agnibha Chandra, Rohit Roy.

As well as playing an essential role in the translation of DNA into proteins, RNA molecules carry out many other essential biological functions in cells, often through their interactions with proteins. A critical challenge in modelling such interactions computationally is that the RNA is often highly flexible, especially in single-stranded (ssRNA) regions of its structure. These flexible regions are often very important because it is through their flexibility that the RNA can adjust its 3D conformation in order to bind to a protein surface. However, conventional protein-protein docking algorithms generally assume that the 3D structures to be docked are rigid, and so are not suitable for modeling protein-RNA interactions. There is therefore much interest in developing protein-RNA docking algorithms which can take RNA flexibility into account.

We are currently developing a novel flexible docking algorithm which first docks small fragments of ssRNA (typically three nucleotides at a time) onto a protein surface, and then combinatorially reassembles those fragments in order to recover a contiguous ssRNA structure on the protein surface [44], [45]. We have since implemented a prototype "forward-backward" dynamic programming algorithm with stochastic backtracking that allows us to model protein RNA interactions for ssRNAs of up to 7 nucleotides without requiring any prior knowledge of the interaction, while still avoiding a brute-force search. In the frame of our PEPS collaboration "InterANRIL" with the IMoPA lab (University of Lorraine), we are currently working with biologists to apply the approach to modeling certain long non-coding RNA (lncRNA) complexes. In order to extend this approach to partially structured RNA molecules, we have built an automated pipeline to create (i) libraries of RNA fragments with arbitrary characteristics such as secondary structure, and (ii) testing benchmarks for applying those libraries in docking. In the frame of our LUE-FEDER CITRAM project we adapted this approach and this pipeline to DNA docking in order to model the complex formed by a bacterial relaxase and its target DNA.

## DYLISS Project-Team

# 4. Application Domains

## 4.1. Application fields in biology

In terms of transfer and societal impact, we consider that our role is to develop fruitful collaborations with laboratories of biology in order to consolidate their studies by a smart use of our tools and prototypes and generate new biological hypotheses to be tested experimentally.

**Marine Biology: seaweed enzymes and metabolism** Our main field of field **marine biology**, in close collaborations with the Roscoff Biological Station, in the framework of the Idealg project. Our goal is to apply our methods based on combinatorial optimization to the reconstruction of genome-scale metabolic networks, the understanding of microbial consortia, and classification of enzyme families. A main application model is *E. Siliculosus*, for which we reconstructed a metabolic network, predicted *HAD* proteins, and suggested new annotations of 56 genes based on metabolic network considerations. These approaches also shed light on evolution of metabolic processes. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpus* and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. For instance, our analyses suggest that the bacterium *Ca. P. ectocarpus* is able to provide both  $\beta$ -alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway. These studies are now extended to the understanding of full host-microbial interactions.

**Micro-biology: elucidating the functioning of extremophile consortiums of bacteria** In this application field, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile. In order to elucidate the main characteristics of these bacteria. In particular, genome-scale metabolic network have been reconstructed for bacteria involved in biomining processes and in Salmon pathogenicity, already leading to a better understanding of bacterial interactions and growth.

**Agriculture and environmental sciences: upstream controllers of pork and pea-aphid metabolism and regulation** In this application field, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues, in collaboration with the INRA centers of Rennes (Pegase, Igepp, Scribe). This is a relevant application field for our researches work on the identification of upstream regulators within large-scale knowledge databases and on semantic-based analysis of metabolic networks, in order to interpreting differences of gene expression in pork meat and figure out the main gene-regulators of the response of porks to several diets, or to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants.

**Health: deciphering pathways involved in the TGF- $\beta$  signalling network** This topic is studied with the IRSET laboratory of Rennes. TGF- $\beta$  is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis with both beneficial anti-tumor effect and pro-tumor effect. Deciphering protumor versus antitumor signaling requires to take into account a system-wide view and develop predictive models for therapeutic benefit. We are developing *Cadbiom* in order to identify gene networks associated with innate immune response to viral infection that combine TGF- $\beta$  and interleukine signaling pathways.

## **ERABLE Project-Team**

### **4. Application Domains**

#### **4.1. Biology and Health**

The main areas of application of ERABLE are: (1) biology understood in its more general sense, with a special focus on symbiosis and on intracellular interactions, and (2) health with a special emphasis for now on infectious diseases, rare diseases, and cancer.

## GENSCALE Project-Team

# 4. Application Domains

## 4.1. Introduction

Today, sequencing data are intensively used in many life science projects. The methodologies developed by the GenScale group are generic approaches that can be applied to a large panel of domains such as health, agronomy or environment areas. The next sections briefly describe examples of our activity in these different domains.

## 4.2. Health

**Genetic and cancer disease diagnostic:** Genetic diseases are caused by some particular mutations in the genomes that alter important cell processes. Similarly, cancer comes from changes in the DNA molecules that alter cell behavior, causing uncontrollable growth and malignancy. Pointing out genes with mutations helps in identifying the disease and in prescribing the right drug. Thus, DNA from individual patients is sequenced and the aim is to detect potential mutations that may be linked to the patient disease. Today the bioinformatics analysis is mainly based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. Tomorrow, due to the decreasing cost of the sequencing process, bioinformatics analysis will scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with cancers.

**Neurodegenerative disorders:** The biological processes that lead from abnormal protein accumulation to neuronal loss and cognitive dysfunction is not fully understood. In this context, neuroimaging biomarkers and statistical methods to study large datasets play a pivotal role to better understand the pathophysiology of neurodegenerative disorders. The discovery of new anatomical biomarkers could thus have a major impact on clinical trials by allowing inclusion of patients at a very early stage, at which treatments are the most likely to be effective. Correlations with genetic variables can determine subgroups of patients with common anatomical and genetic characteristics.

## 4.3. Agronomy and Environment

**Improving plant breeding:** such projects aim at 1) identifying favorable alleles at loci contributing to phenotypic variation, 2) characterizing polymorphism at the functional level and 3) providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

**Insect genomics:** Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities [14].

**Ocean biodiversity:** The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and their role, for example, in the CO<sub>2</sub> sequestration.

**IBIS Project-Team (section vide)**

## LIFEWARE Project-Team

### 4. Application Domains

#### 4.1. Preamble

Our collaborative work on biological applications is expected to serve as a basis for groundbreaking advances in cell functioning understanding, cell monitoring and control, and novel therapy design and optimization. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Furthermore, one important application of our research is the development of a **modeling software** for computational systems biology.

#### 4.2. Modeling software for systems biology and synthetic biology

Since 2002, we develop an open-source software environment for modeling and analyzing biochemical reaction systems. This software, called the Biochemical Abstract Machine (**BIOCHAM**), is compatible with SBML for importing and exporting models from repositories such as BioModels. It can perform a variety of static analyses, specify behaviors in Boolean or quantitative temporal logics, search parameter values satisfying temporal constraints, and make various simulations. While the primary reason of this development effort is to be able to **implement our ideas and experiment them quickly on a large scale**, BIOCHAM is used by other groups either for building models, for comparing techniques, or for teaching (see statistics in software section). BIOCHAM-WEB is a web application which makes it possible to use BIOCHAM without any installation. We plan to continue developing BIOCHAM for these different purposes and improve the software quality.

#### 4.3. Coupled models of the cell cycle and the circadian clock

Recent advances in cancer chronotherapy techniques support the evidence that there exist important links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characteristics. These questions are at the heart of our collaboration with Franck Delaunay (CNRS Nice) and Francis Lévi (Univ. Warwick, GB, formerly INSERM Hopital Paul Brousse, Villejuif) and of our participation in the ANR **HYCLOCK** project and in the submitted EU H2020 C2SyM proposal, following the former EU EraNet Sysbio **C5Sys** and FP6 **TEMPO** projects. In the past, we developed a coupled model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints<sup>0</sup>. We now focus on the bidirectional coupling between the cell cycle and the circadian clock and expect to gain fundamental insights on this complex coupling from computational modeling and single-cell experiments.

#### 4.4. Biosensor design and implementation in non-living protocells

In collaboration with Franck Molina (CNRS, Sys2Diag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan) we ambition to apply our techniques to the design and implementation of biosensors in non-living vesicles for medical applications. Our approach is based on purely protein computation and on our ability to compile controllers and programs in biochemical reactions. The realization will be prototyped using a microfluidic device at CNRS Sys2Diag which will allow us to precisely control the size of the vesicles and the concentrations of the injected proteins. It is worth noting that the choice of non-living chassis, in contrast to living cells in synthetic biology, is particularly appealing for security considerations and compliance to forthcoming EU regulation.

---

<sup>0</sup>Elisabetta De Maria, François Fages, Aurélien Rizk, Sylvain Soliman. Design, Optimization, and Predictions of a Coupled Model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints. Theoretical Computer Science, 412(21):2108-2127, 2011.

## **4.5. Functional characterization of the resistance of bacterial populations to antimicrobial treatments**

Antibiotic resistance is becoming a problem of central importance at a global level. Two mechanisms are at the origin of non-susceptibility to antimicrobial treatments. The first one comes from adaptation of bacterial cells to antibacterial treatments, notably through the modification of efflux pumps or the expression of enzymes that degrade the antibiotics. Cells are individually resistant. The second one, typically found in resistances to  $\beta$ -lactams, a broad class of antibiotics, originates from the release in the environment of the antibiotic degrading enzymes by the dead cells. This leads to population effects by which cells become collectively resilient.

The functional characterization of these different effects is important for the best use of antibiotics (antibiotic stewardship). In collaboration with Lingchong You (Duke University) and with Philippe Glaser (Institut Pasteur), we develop experimental platforms, models, and optimal model calibration methods that gives precise estimations of individual resistance and collective resilience of bacterial populations to antibiotic treatments.



**MORPHEME Project-Team (section vide)**

**MOSAIC Team (section vide)**

## PLEIADE Team

# 4. Application Domains

## 4.1. Genome and transcriptome annotation, to model function

Sequencing genomes and transcriptomes provides a picture of how a biological system can function, or does function under a given physiological condition. Simultaneous sequencing of a group of related organisms is now a routine procedure in biological laboratories for studying a behavior of interest, and provides a marvelous opportunity for building a comprehensive knowledge base of the relations between genomes. Key elements in mining these relations are: classifying the genes in related organisms and the reactions in their metabolic networks, recognizing the patterns that describe shared features, and highlighting specific differences.

PLEIADE will develop applications in comparative genomics of related organisms, using new mathematical tools for representing compactly, at different scales of difference, comparisons between related genomes. New methods based on computational geometry refine these comparisons. Compact representations can be stored, exchanged, and combined. They will form the basis of new simultaneous genome annotation methods, linked directly to abductive inference methods for building functional models of the organisms and their communities.

Our ambition in biotechnology is to permit the design of synthetic or genetically selected organisms at an abstract level, and guide the modification or assembly of a new genome. Our effort is focused on two main applications: genetic engineering and synthetic biology of oil-producing organisms (biofuels in CAER, palm oils), and improving and selecting starter microorganisms used in winemaking (collaboration with the ISVV and the BioLaffort company).

### 4.1.1. Oil Palm lipid synthesis

The largest source of vegetable oil<sup>0</sup> is the fruit mesocarp of the oil palm *Elaeis guineensis*, a remarkable tissue that can accumulate up to 90% oil, the highest level observed in the plant kingdom. The market share of oil palm is expected to increase in order to meet increased demand for vegetable oil, predicted to double by 2030 [18], be it as food or as a source of biofuels in Africa. A significant proportion of palm oil is produced on small estates that do not have access to efficient milling facilities, and run a great risk of spoilage through oil acidification. Improving palm oil quality through genetics and selection will result in economic gains [24] by addressing several targets such as improvement of oil yield, tuning of oil quality through the rate of unsaturated fatty acids or impairment of degradation processes. Furthermore, as genome biodiversity resides mostly in Africa, oil from African oil palms can vary greatly in fatty acid composition according to cultivar genetic differences and to weather conditions, and the precise mechanisms regulating this variability are not yet understood.

A growing body of molecular resources for studying oil palm fruit are making it possible to study and improve the quality and quantity of oil produced by oil palms. In particular, these oils can vary greatly in fatty acid composition, and while the precise mechanisms regulating this variability are not completely understood, establishing a link between oil palm genotype and phenotype appears increasingly feasible. PLEIADE will work with the CNRS/UB UMR 5200 (LBM), a laboratory with an established reputation in studying fatty acid metabolism in *E. guineensis*, to improve understanding of the links between genetic diversity and oil production, and participate in developing applications.

### 4.1.2. Engineering pico-algae

---

<sup>0</sup>32% of the world market share [24]

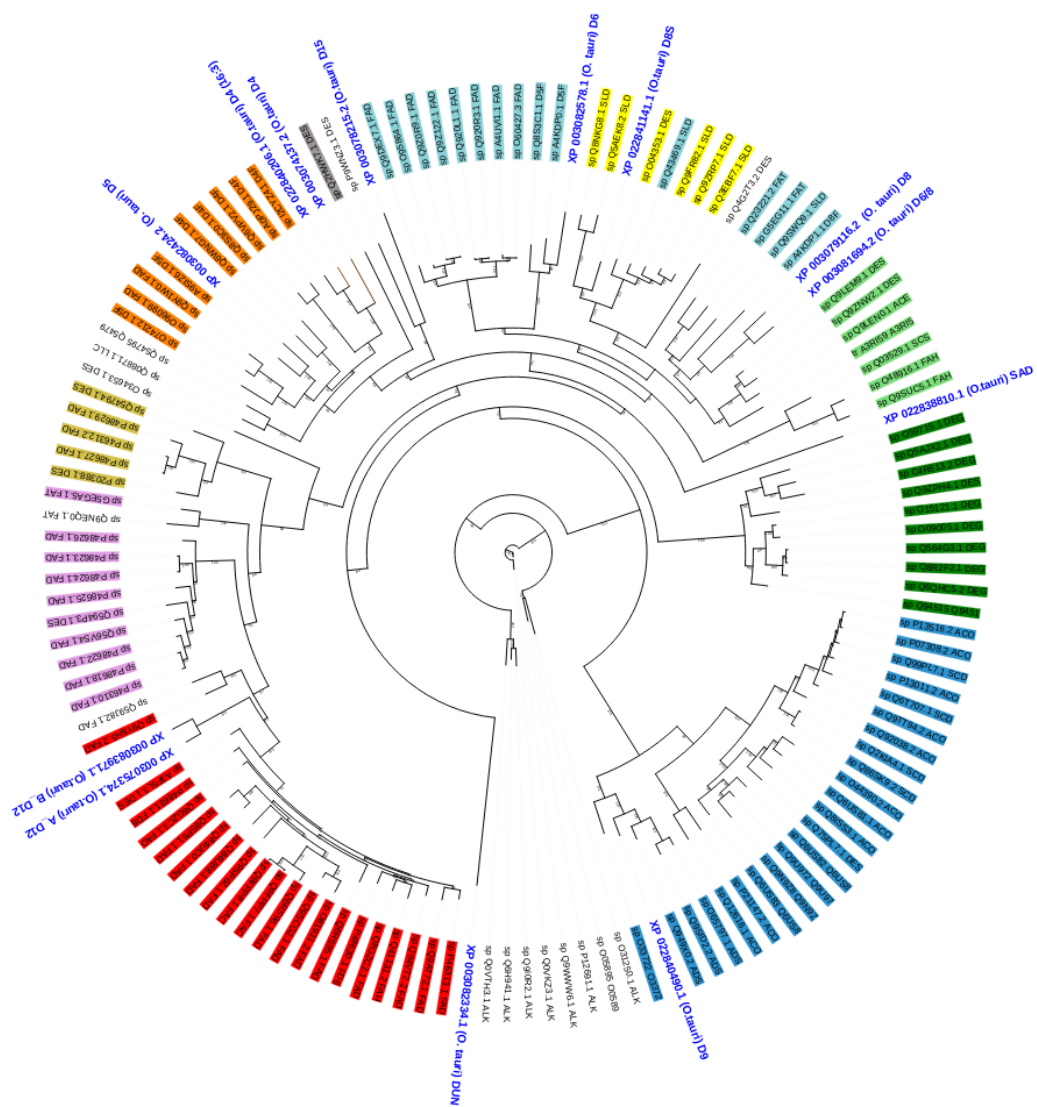


Figure 3. Phylogenetic structure of long-chain polyunsaturated fatty acid desaturase specificity. Highlighted are thirteen desaturases from *Ostreococcus tauri*

Docosahexaenoic acid (DHA) is an essential nutriment for human brain tissue and can only be obtained from marine or riverine fish that live on phytoplankton and zooplankton, since human neurons lack the delta desaturase required for *de novo* synthesis of DHA. [21] Unfortunately, fishing is become less and less a sustainable resource. Since phytoplankton and zooplankton are the ultimate source of DHA consumed, there is considerable interest in obtaining DHA directly rather than through the intermediary of fish. A very promising approach is through the bio-engineering of pico-algae.

In order to produce the long-chain polyunsaturated fatty acids (LC-PUFA) needed for human nutrition, it is necessary to precisely engineer the desaturases that produce them. Desaturases are enzymes responsible for the introduction of double bonds into fatty acids. Desaturases are specific in recognizing their substrates and in placing the double bond in the proper place. The desaturases that produce the LC-PUFA necessary for human nutrition are present only in some species.

Our goal is to design methods to predict the substrate and region specificities for desaturases in algal species, particularly *Ostreococcus tauri*, the smallest photosynthetic eukaryote that can be cultivated. Thirteen desaturases are known in *O. tauri* and can be placed in the phylogeny of the desaturase family (figure 3). The biochemical and structural characterization of these enzymes is as yet very incomplete. This work is ongoing (see section 8.2.3) and requires close collaboration between biologists and computer scientists.

## 4.2. Molecular based systematics and taxonomy

Defining and recognizing myriads of species in biosphere has taken phenomenal energy over the past centuries and remains a major goal of Natural History. It is an iconic paradigm in pattern recognition (clustering has coevolved with numerical taxonomy many decades ago). Developments in evolution and molecular biology, as well as in data analysis, have over the past decades enabled a profound revolution, where species can be delimited and recognized by data analysis of sequences. We aim at proposing new tools, in the framework of E-science, which make possible (i) better exploration of the diversity in a given clade, and (ii) assignment of a place in these patterns for new, unknown organisms, using information provided by sets of sequences. This will require investment in data analysis, machine learning, and pattern recognition to deal with the volumes of data and their complexity.

One example of this project is about the diversity of trees in Amazonian forest, in collaboration with botanists in French Guiana. Protists (unicellular Eukaryotes) are by far more diverse than plants, and far less known. Molecular exploration of Eukaryotes diversity is nowadays a standard in biodiversity studies. Data are available, through metagenomics, as an avalanche and make molecular diversity enter the domain of Big Data. Hence, an effort will be invested, in collaboration with other Inria teams (GenScale, HiePACS) for porting to HPC algorithms of pattern recognition and machine learning, or distance geometry, for these tools to be available as well in metagenomics. This will be developed first on diatoms (unicellular algae) in collaboration with INRA team at Thonon and University of Uppsala), on pathogens of tomato and grapewine, within an existing network, and on bacterial communities, in collaboration with University of Pau. For the latter, the studies will extend to correlations between molecular diversity and sets of traits and functions in the ecosystem.

## 4.3. Community ecology and population genetics

Community assembly models how species can assemble or disassemble to build stable or metastable communities. It has grown out of inventories of countable organisms. Using *metagenomics* one can produce molecular based inventories at rates never reached before. Most communities can be understood as pathways of carbon exchange, mostly in the form of sugar, between species. Even a plant cannot exist without carbon exchange with its rhizosphere. Two main routes for carbon exchange have been recognized: predation and parasitism. In predation, interactions—even if sometimes dramatic—may be loose and infrequent, whereas parasitism requires what Claude Combes has called intimate and sustainable interactions [19]. About one decade ago, some works [25] have proposed a comprehensive framework to link the studies of biodiversity with community assembly. This is still incipient research, connecting community ecology and biogeography.

We aim at developing graph-based models of co-occurrence between species from NGS inventories in metagenomics, i.e. recognition of patterns in community assembly, and as a further layer to study links, if any, between diversity at different scales and community assemblies, starting from current, but oversimplified theories, where species assemble from a regional pool either randomly, as in neutral models, or by environmental filtering, as in niche modeling. We propose to study community assembly as a multiscale process between nested pools, both in tree communities in Amazonia, and diatom communities in freshwaters. This will be a step towards community genomics, which adds an ecological flavour to metagenomics.

Convergence between the processes that shape genetic diversity and community diversity—drift, selection, mutation/speciation and migration—has been noted for decades and is now a paradigm, establishing a continuous scale between levels of diversity patterns, beyond classical approaches based on iconic levels like species and populations. We will aim at deciphering diversity pattern along these gradients, connecting population and community genetics. Therefore, some key points must be addressed on reliability of tools.

Next-generation sequencing technologies are now an essential tool in population and community genomics, either for making evolutionary inferences or for developing SNPs for population genotyping analyses. Two problems are highlighted in the literature related to the use of those technologies for population genomics: variable sequence coverage and higher sequencing error in comparison to the Sanger sequencing technology. Methods are developed to develop unbiased estimates of key parameters, especially integrating sequencing errors [23]. An additional problem can be created when sequences are mapped on a reference sequence, either the sequenced species or an heterologous one, since paralogous genes are then considered to be the same physical position, creating a false signal of diversity [20]. Several approaches were proposed to correct for paralogy, either by working directly on the sequences issued from mapped reads [20] or by filtering detected SNPs. Finally, an increasingly popular method (RADseq) is used to develop SNP markers, but it was shown that using RADseq data to estimate diversity directly biases estimates [12]. Workflows to implement statistical methods that correct for diversity biases estimates now need an implementation for biologists.

## SERPICO Project-Team

### 4. Application Domains

#### 4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale

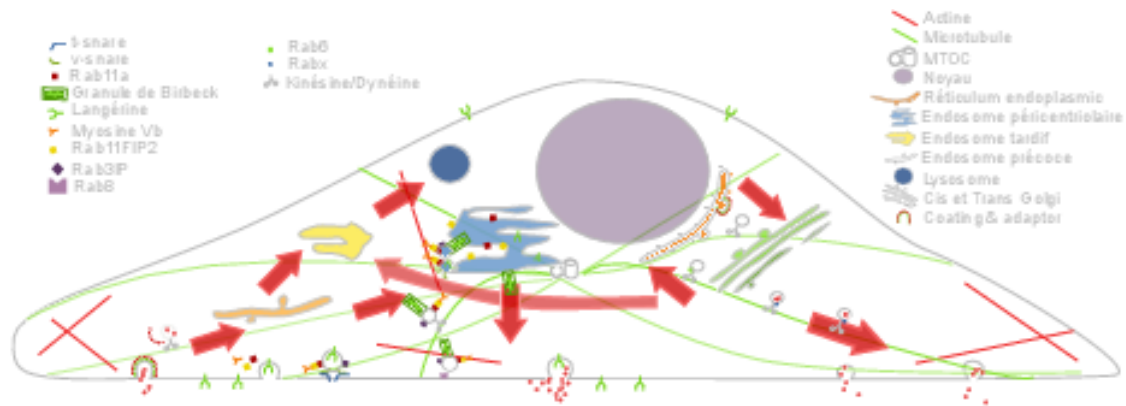


Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried at UMR 144 CNRS-Institut Curie (“Space Time imaging of Endomembranes and organelles Dynamics” team) contributed to a better understanding of the intracellular compartmentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight: i/ multiple sorting and structural events involved in the biogenesis of these organelles; ii/ complexity of the endo-melanosomal network of these highly specialized cells; iii/ complex molecular architecture organizing and coordinating their dynamics; iv/ intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;
2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Fig. 1 ). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and essential issues in this context.

## 4.2. Imaging and analysis of cytoskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells in blood and organs and eventually the formation of secondary tumors and metastases.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

- i/ the actin filaments promote cell protrusion, adhesion and retraction;
- ii/ the microtubules are the support of molecule traffic and cell polarization;
- iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and molecule trafficking.



## ARAMIS Project-Team

# 4. Application Domains

## 4.1. Introduction

We develop different applications of our new methodologies to brain pathologies, mainly neurodegenerative diseases. These applications aim at:

- better understanding the pathophysiology of brain disorders;
- designing systems to support clinical decisions such as diagnosis, prognosis and design of clinical trials;
- developing brain computer interfaces for clinical applications.

## 4.2. Understanding brain disorders

Computational and statistical approaches have the potential to help understand the pathophysiology of brain disorders. We first aim to contribute to better understand the relationships between pathological processes, anatomical and functional alterations, and symptoms. Moreover, within a single disease, there is an important variability between patients. The models that we develop have the potential to identify more homogeneous disease subtypes, that would constitute more adequate targets for new treatments. Finally, we aim to establish the chronology of the different types of alterations. We focus these activities on neurodegenerative diseases: dementia (Alzheimer's disease, fronto-temporal dementia), Parkinson's disease, multiple sclerosis.

## 4.3. Supporting clinical decisions

We aim to design computational tools to support clinical decisions, including diagnosis, prognosis and the design of clinical trials. The differential diagnosis of neurodegenerative diseases can be difficult. Our tools have the potential to help clinicians by providing automated classification that can integrate multiple types of data (clinical/cognitive tests, imaging, biomarkers). Predicting the evolution of disease in individual patients is even more difficult. We aim to develop approaches that can predict which alterations and symptoms will occur and when. Finally, new approaches are needed to select participants in clinical trials. Indeed, it is widely recognized that, to have a chance to be successful, treatments should be administered at a very early stage.

## 4.4. Brain computer interfaces for clinical applications

A brain computer interface (BCI) is a device aiming to decode brain activity, thus creating an alternate communication channel between a person and the external environment. BCI systems can be categorized on the basis of the classification of an induced or evoked brain activity. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Hence adaptation, as well as learning, is a key component of a BCI because users must learn to modulate their brainwaves to generate distinct brain patterns. Usually, a BCI is considered a technology for people to substitute some lost functions. However, a BCI could also help in clinical rehabilitation to recover motor functions. Indeed, in current neuroscience-based rehabilitation it is recognized that protocols based on mental rehearsal of movements (like motor imagery practicing) are a way to access the motor system because they can induce an activation of sensorimotor networks that were affected by lesions. Hence, a BCI based on movement imagery can objectively monitor patients' progress and their compliance with the protocol, monitoring that they are actually imagining movements. It also follows that feedback from such a BCI can provide patients with an early reinforcement in the critical phase when there is not yet an overt sign of movement recovery.

## ATHENA Project-Team

# 4. Application Domains

## 4.1. Applications of diffusion MRI

### **Clinical domain: Diagnosis of neurological disorder**

Various examples of CNS diseases as Alzheimer's and Parkinson's diseases and others like multiple sclerosis, traumatic brain injury and schizophrenia have characteristic abnormalities in the micro-structure of brain tissues that are not apparent and cannot be revealed reliably by standard imaging techniques. Diffusion MRI can make visible these co-lateral damages to the fibers of the CNS white matter that connect different brain regions. This is why in our research, Diffusion MRI is the structural imaging modality that will be considered to recover the CNS connectivity.

## 4.2. Applications of M/EEG

### **Clinical domain: Diagnosis of neurological disorders**

The dream of all M/EEG researchers is to alleviate the need for invasive recordings (electrocorticograms or intracerebral electrodes), which are often necessary prior to brain surgery, in order to precisely locate both pathological and vital functional areas. We are involved in this quest, particularly through our collaborations with the La Timone hospital in Marseille.

Subtopics include:

- Diagnosis of neurological disorders such as epilepsy, schizophrenia, tinnitus, ...
- Presurgical planning of brain surgery.
- Collaboration with the *Institut de Neurosciences des Systèmes* on these topics <http://ins.univ-amu.fr/>

### **Cognitive research**

- Aims at better understanding the brain spatio-temporal organisation.
- Collaboration with the *Laboratory for Neurobiology of Cognition* in order to develop methods that suit their needs for sophisticated data analysis.

**Brain Computer Interfaces (BCI)** aim to allow direct control of external devices using brain signals such as measured through EEG. In our project, BCI can be seen as an application of EEG processing techniques, but also as an object of fundamental and applied research as they open the way for more dynamical and active brain cognitive protocols.

We develop a research collaboration with the eemagine/ANT-Neuro company. We collaborate with Nice University Hospital on the usage of BCI-based communication for ALS<sup>0</sup> patients.

---

<sup>0</sup>Nice University Hospital hosts a regional reference center for patients suffering from Amyotrophic Lateral Sclerosis

## BIOVISION Project-Team

### 4. Application Domains

#### 4.1. Applications of virtual/augmented reality for low-vision

- **Rehabilitation:** Serious games are games designed for a primary purpose which is not pure entertainment. In our context, we think about serious games as a way to help low-vision patients in performing rehabilitation exercises. Virtual/augmented reality technology is a promising platform to develop such rehabilitation exercises targeted to specific pathologies. For example, with Age Macular Degeneration (AMD), our objective is to propose solutions allowing rehabilitation of visuo-perceptual-motor functions to optimally use residual portions of the peripheral retina and obtain efficient “eccentric viewing”.
- **Vision aid-systems:** A variety of aids for low-vision people are already on the market using different kinds of virtual/augmented reality platforms (dedicated or large public ones). They offer different functionalities (magnification, image enhancement, text to speech, face and object recognition). Our goal is to design new solutions allowing autonomous interaction in mixed reality environments, and take advantage of the improvement of functions obtained via rehabilitation protocols.
- **Cognitive research:** Virtual/augmented reality technology represents a new opportunity to conduct cognitive and behavioural research using virtual environments where all parameters can be psychophysically controlled. Our objective is to re-assess common theories by allowing patients to freely explore their environment in more ecological conditions.

#### 4.2. Applications of vision modeling studies

- **Neuroscience research.** Making in-silico experiments is a way to reduce the experimental costs, to test hypotheses and design models, and to test algorithms. Our goal is to develop a large-scale simulations platform of impaired retinas, called Macular, allowing to mimic specific degeneracies or pharmacologically induced impairments, as well as to emulate electric stimulation by prostheses. In addition, the platform provides a realistic entry to models or simulators of the thalamus or the visual cortex, in contrast to the entries usually considered in modelling studies.
- **Education.** Macular is also targeted as a useful tool for educational purposes, illustrating for students how the retina works and respond to visual stimuli.

**CAMIN Team (section vide)**

**EPIONE Project-Team (section vide)**

## GALEN-POST Team

### 4. Application Domains

#### 4.1. Representation Learning for Network Biology

**Participants:** Fragkiskos Malliaros, Abdulkadir Çelikkanat (Collaboration: Duong Nguyen, UC San Diego)  
Networks (or graphs) are ubiquitous in the domain of biology, as many biological systems can naturally be mapped to graph structures. Characteristic examples include protein-protein interaction and gene regulatory networks. To this extend, machine learning on graphs is an important task with many practical applications in network biology. For example, in the case on protein-protein interaction networks, predicting the function of a protein is a key task that assigns biochemical roles to proteins. The main challenge here is to find appropriate representations of the graph structure, in order to be easily exploited by machine learning models. The traditional approach to the problem was relying on the extraction of “hand-crafted” discriminating features that encode information about the graph, based on user-defined heuristics. Nevertheless, this approach has demonstrated severe limitations, as the learning process heavily depends on the manually extracted features. To this end, feature (or representation) learning techniques can be used to automatically learn to encode the graph structure into low-dimensional feature vectors – which can later be used in learning tasks. Our goal here is to develop a systematic framework for large-scale representation learning on biological graphs. Our approach takes advantage of the clustering structure of these networks, to further enhance the ability of the learned features to capture intrinsic structural properties.

#### 4.2. Breast tomosynthesis

**Participants:** Emilie Chouzenoux, Jean-Christophe Pesquet, Maissa Sghaier (collaboration G. Palma, GE Healthcare)

Breast cancer is the most frequently diagnosed cancer for women. Mammography is the most used imagery tool for detecting and diagnosing this type of cancer. Since it consists of a 2D projection method, this technique is sensitive to geometrical limitations such as the superimposition of tissues which may reduce the visibility of lesions or make even appear false structures which are interpreted by radiologists as suspicious signs. Digital breast tomosynthesis allows these limitations to be circumvented. This technique is grounded on the acquisition of a set of projections with a limited angle view. Then, a 3D estimation of the sensed object is performed from this set of projections, so reducing the overlap of structures and improving the visibility and detectability of lesions possibly present in the breast. The objective of our work is to develop a high quality reconstruction methodology where the full pipeline of data processing will be modeled.

#### 4.3. Inference of gene regulatory networks

**Participants:** Jean-Christophe Pesquet (collaboration A. Pirayre and L. Duval, IFPEN)

The discovery of novel gene regulatory processes improves the understanding of cell phenotypic responses to external stimuli for many biological applications, such as medicine, environment or biotechnologies. To this purpose, transcriptomic data are generated and analyzed from DNA microarrays or more recently RNAseq experiments. They consist in genetic expression level sequences obtained for all genes of a studied organism placed in different living conditions. From these data, gene regulation mechanisms can be recovered by revealing topological links encoded in graphs. In regulatory graphs, nodes correspond to genes. A link between two nodes is identified if a regulation relationship exists between the two corresponding genes. In our work, we propose to address this network inference problem with recently developed techniques pertaining to graph optimization. Given all the pairwise gene regulation information available, we propose to determine the presence of edges in the considered GRN by adopting an energy optimization formulation integrating additional constraints. Either biological (information about gene interactions) or structural (information about node connectivity) a priori are considered to restrict the space of possible solutions. Different priors lead to different properties of the global cost function, for which various optimization strategies, either discrete and continuous, can be applied.

#### 4.4. Imaging biomarkers and characterization for chronic lung diseases

**Participants:** Guillaume Chassagnon, Maria Vakalopoulou (in collaboration with Evangelia Zacharaki and Nikos Paragios: University of Patras; Therapanacea)

Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. Our goal is to develop quantitative imaging biomarkers that allow (i) severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis as models for restrictive and obstructive lung disease, respectively. Two different approaches are investigated: disease assessment by deep convolutional neural networks and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

#### 4.5. Imaging radiomics and genes to assess immunotherapy

**Participants:** Roger Sun, Théo Estienne, Enzo Batistella, Maria Vakalopoulou (in collaboration with Éric Deutsch and Nikos Paragios: Institut de Cancérologie Gustave Roussy, Therapanacea)

Because responses of patients with cancer to immunotherapy can vary in success, innovative predictors of response to treatment are urgently needed to improve treatment outcomes. We aimed to develop and independently validate a radiomics-based biomarker of tumour-infiltrating CD8 cells in patients included in phase 1 trials of anti-programmed cell death protein (PD)-1 or anti-programmed cell death ligand 1 (PD-L1) monotherapy. We also aimed to evaluate the association between the biomarker, and tumour immune phenotype and clinical outcomes of these patients. This work is in collaboration with the Institut de Cancérologie Gustave Roussy Paris.

#### 4.6. Restoration of old video archives

**Participants:** Emilie Chouzenoux, Jean-Christophe Pesquet (collaboration F. Abboud, WITBE, J.-H. Chenot and L. Laborelli, INA)

The last century has witnessed an explosion in the amount of video data stored with holders such as the National Audiovisual Institute whose mission is to preserve and promote the content of French broadcast programs. The cultural impact of these records, their value is increased due to commercial reexploitation through recent visual media. However, the perceived quality of the old data fails to satisfy the current public demand. The purpose of our work is to propose new methods for restoring video sequences supplied from television archive documents, using modern optimization techniques with proven convergence properties [21], [50], [3].

#### 4.7. Development of a heart ventricle vessel generation model for perfusion analysis

**Participant:** Hugues Talbot (collaboration with L. Najman ESIEE Paris, I. Vignon-Clementel, REO Team, Inria, Leo Grady, Heartflow Inc.)

Cardio-vascular diseases are the leading cause of mortality in the world. Understanding these diseases is still a current, challenging and essential research project. The leading cause of heart malfunction are stenoses causing ischemia in the coronary vessels. Current CT and MRI technology allow to assess coronary diseases but are typically invasive, requiring catheterization and relatively toxic contrast agents injection. With Heartflow, a US based company, we have in the past worked to use image-based exams only, limiting the use of contrast agents and in many cases eliminating catheterisation. Heartflow is current the market leader in non-invasive coronary exams.

Unfortunately, current imaging technology is unable to assess the full length of coronary vessels. CT is limited to a resolution of about 1mm, whereas coronary vessels can be much smaller. Blood perfusion throughout the heart muscle can provide insight regarding coronary health in areas that CT or MRI cannot assess. Perfusion imaging with PET or a Gamma camera, the current gold standard, is an invasive technology requiring the use of radioactive tracers.

With heartflow and with the REO team of Inria, we have investigated patient-specific vessel generation models together with porous model simulations in order to propose a forward model of perfusion imaging, based on the known patient data, computer flow dynamic simulations as well as experimental data consistent with known vessel and heart muscle physiology. The objective of this work is to solve the inverse problem of locating and assessing coronary diseases even though the affected vessels are too small to be imaged directly.



**MATHNEURO Team (section vide)**

## MIMESIS Team

# 4. Application Domains

## 4.1. Surgical Training

Virtual training prevents medical students from early manipulation of real patients. The development of simulation used for medical training usually requires important computational power, since realistic behaviours are key to deliver a high-fidelity experience to the trainee. Further, the quality of interaction with the simulator (usually via visual and haptic rendering) is also of paramount importance. All these constraints make the development of training systems time-consuming thus limiting the deployment of virtual simulators in standard medical curriculum.

## 4.2. Pre-operative Planning

Beyond training, clinicians ask for innovative tools that can assist them in the pre-operative planning of an intervention. Using the patient information acquired before the operation, physics-based simulations allow to simulate the effect of therapy with no risk to the patient. The clinicians can thus virtually assess different strategies and select the optimal procedure. Compared to a training simulation, a planning system requires a high accuracy to ensure reliability. Constrained by the time elapsed between the preoperative acquisition and the intervention, the computation must also be efficient.

## 4.3. Intra-operative Navigation

Besides the surgery training and planning, another major need from clinicians is surgical guidance. While the practitioner is performing the operation, a guidance system provides enriched visual feedback. This is especially useful with the emergence of minimally invasive surgery (MIS) where the visual information is often strongly limited. It can be used for example to avoid critical areas such as vessels or to highlight the position of a tumour during its resection. In the MIS technique, the clinician does not interact with organs directly as in the open surgery, but manipulates instruments inserted through trocars placed in small incisions in the wall of the abdominal cavity. The surgeon can observe these instruments on a display showing a video stream captured by an endoscopic camera inserted through the navel. The main advantage of the method resides in reducing pain and time recovery, in addition to reducing bleeding and risks of infection. However, from a surgical standpoint, the procedure is quite complex since the field of view is considerably reduced and the direct manipulation of organs is not possible.

## **MNEMOSYNE Project-Team**

# **4. Application Domains**

## **4.1. Overview**

One of the most original specificities of our team is that it is part of a laboratory in Neuroscience (with a large spectrum of activity from the molecule to the behavior), focused on neurodegenerative diseases and consequently working in tight collaboration with the medical domain. As a consequence, neuroscientists and the medical world are considered as the primary end-users of our researches. Beyond data and signal analysis where our expertise in machine learning may be possibly useful, our interactions are mainly centered on the exploitation of our models. They will be classically regarded as a way to validate biological assumptions and to generate new hypotheses to be investigated in the living. Our macroscopic models and their implementation in autonomous robots will allow an analysis at the behavioral level and will propose a systemic framework, the interpretation of which will meet aetiological analysis in the medical domain and interpretation of intelligent behavior in cognitive neuroscience.

The study of neurodegenerative diseases is targeted because they match the phenomena we model. Particularly, the Parkinson disease results from the death of dopaminergic cells in the basal ganglia, one of the main systems that we are modeling. The Alzheimer disease also results from the loss of neurons, in several cortical and extracortical regions. The variety of these regions, together with large mnemonic and cognitive deficits, require a systemic view of the cerebral architecture and associated functions, very consistent with our approach.

Of course, numerical sciences are also impacted by our researches, at several levels. At a global level, we will propose new control architectures aimed at providing a higher degree of autonomy to robots, as well as machine learning algorithms working in more realistic environment. More specifically, our focus on some cognitive functions in closed loop with a real environment will address currently open problems. This is obviously the case for planning and decision making; this is particularly the case for the domain of affective computing, since motivational characteristics arising from the design of an artificial physiology allow to consider not only cold rational cognition but also hot emotional cognition. The association of both kinds of cognition is undoubtedly an innovative way to create more realistic intelligent systems but also to elaborate more natural interfaces between these systems and human users.

At last, we think that our activities in well-founded distributed computations and high performance computing are not just intended to help us design large scale systems. We also think that we are working here at the core of informatics and, accordingly, that we could transfer some fundamental results in this domain.

## NEUROSYS Project-Team

# 4. Application Domains

## 4.1. Medical applications

Our research directions are motivated by applications with a high healthcare or social impact. They are developed in collaboration with medical partners, neuroscientists and psychologists. Almost all of our applications can be seen as neural interfaces which require analysis and modeling of sensorimotor rhythms.

### 4.1.1. *Per-operative awareness during general anesthesia*

*Collaborators:* Univ. Hospital of Nancy-Brabois/dept. Anesthesia & Resuscitation

During general anesthesia, brain oscillations change according to the anesthetic drug concentration. Nowadays, 0.2 to 1.3% of patients regain consciousness during surgery and suffer from post-traumatic disorders. Despite the absence of subject movements due to curare, an electroencephalographic analysis of sensorimotor rhythms can help to detect an intention of movement. Within a clinical protocol, we are working on a brain-computer interface adapted to the detection of intraoperative awareness.

### 4.1.2. *Recovery after stroke*

*Collaborators:* Regional Institute of Physical Medicine and Rehabilitation/Center for Physical Medicine and Rehabilitation (Lay St Christophe), Univ. of Lorraine/PERSEUs.

Stroke is the main cause of acquired disability in adults. Neurosys aims at recovering limb control by improving the kinesthetic motor imagery (KMI) generation of post-stroke patients. We propose to design a KMI-based EEG neural interface which integrates complementary modalities of interactions such as tangible and haptic ones to stimulate the sensorimotor loop. This solution would provide a more engaging and compelling stroke rehabilitation training program based on KMI production.

### 4.1.3. *Modeling Parkinson's disease*

*Collaborators:* Center for Systems Biomedicine (Luxembourg), Institute of Neurodegenerative Diseases (Bordeaux), Human Performance & Robotics laboratory (California State Univ., Long Beach).

Effective treatment of Parkinson's disease should be based on a realistic model of the disease. We are currently developing a neuronal model based on Hodgkin-Huxley neurons reproducing to a certain extent the pathological synchronization observed in basal ganglia in Parkinsonian rats. Moreover, our mesoscopic models of plastic CPG neural circuitries involved in rhythmic movements will allow us to reproduce incoherent coordination of limbs observed on humans affected by Parkinson's diseases like frozen gait, crouch gait. Our long-term objective is to understand how oscillatory activity in the basal ganglia affects motor control in spinal structures.

## PARIETAL Project-Team

# 4. Application Domains

## 4.1. Cognitive neuroscience

### 4.1.1. Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electrophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain, such as: How is the function of cortical territories constrained by anatomy (connectivity)? How to assess the specificity of brain regions? How can one characterize reliably inter-subject differences?

### 4.1.2. Analysis of brain Connectivity

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

### 4.1.3. Modeling of brain processes (MEG)

While fMRI has been very useful in defining the function of regions at the mm scale, Magnetoencephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

### 4.1.4. Current challenges in human neuroimaging (acquisition+analysis)

Human neuroimaging targets two major goals: *i*) the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii*) the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial<sup>0</sup> resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.
- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

---

<sup>0</sup>and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements<sup>0</sup> and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [70], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [71]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as **NeuroVault** or **OpenfMRI**. A framework for integrating such datasets is however still missing.

---

<sup>0</sup>The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

## VISAGES Project-Team

### 4. Application Domains

#### 4.1. Neuroimaging

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions. Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

#### 4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define *in vivo* those four different MS lesion patterns. An *in vivo* distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types. However, it has a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and gray matter. Our major objective within this application domain is then to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g., nD+t MRI) in the brain and the spinal cord. In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labeling neuroimaging (e.g., MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g., USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved *in vivo* characterization of MS pathology.

#### 4.3. Modeling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field, and by continuing our local and wider collaborations. An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g., variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries, etc.). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

## AIRSEA Project-Team

# 4. Application Domains

## 4.1. The Ocean-Atmosphere System

The evolution of natural systems, in the short, mid, or long term, has extremely important consequences for both the global Earth system and humanity. Forecasting this evolution is thus a major challenge from the scientific, economic, and human viewpoints.

Humanity has to face the problem of **global warming**, brought on by the emission of greenhouse gases from human activities. This warming will probably cause huge changes at global and regional scales, in terms of climate, vegetation and biodiversity, with major consequences for local populations. Research has therefore been conducted over the past 15 to 20 years in an effort to model the Earth's climate and forecast its evolution in the 21st century in response to anthropic action.

With regard to short-term forecasts, the best and oldest example is of course **weather forecasting**. Meteorological services have been providing daily short-term forecasts for several decades which are of crucial importance for numerous human activities.

Numerous other problems can also be mentioned, like **seasonal weather forecasting** (to enable powerful phenomena like an El Niño event or a drought period to be anticipated a few months in advance), **operational oceanography** (short-term forecasts of the evolution of the ocean system to provide services for the fishing industry, ship routing, defense, or the fight against marine pollution) or the prediction of **floods**.

As mentioned previously, mathematical and numerical tools are omnipresent and play a fundamental role in these areas of research. In this context, the vocation of AIRSEA is not to carry out numerical prediction, but to address mathematical issues raised by the development of prediction systems for these application fields, in close collaboration with geophysicists.



## ANGE Project-Team

# 4. Application Domains

## 4.1. Overview

Sustainable development and environment preservation have a growing importance and scientists have to address difficult issues such as: management of water resources, renewable energy production, bio/geo-chemistry of oceans, resilience of society w.r.t. hazardous flows, urban pollutions, ...

As mentioned above, the main issue is to propose models of reduced complexity, suitable for scientific computing and endowed with stability properties (continuous and/or discrete). In addition, models and their numerical approximations have to be confronted with experimental data, as analytical solutions are hardly accessible for these problems/models. A. Mangeney (IPGP) and N. Goutal (EDF) may provide useful data.

## 4.2. Geophysical flows

Reduced models like the shallow water equations are particularly well-adapted to the modelling of geophysical flows since they are characterized by large time or/and space scales. For long time simulations, the preservation of equilibria is essential as global solutions are a perturbation around them. The analysis and the numerical preservation of non-trivial equilibria, more precisely when the velocity does not vanish, are still a challenge. In the fields of oceanography and meteorology, the numerical preservation of the so-called geostrophic state, which is the balance between the gravity field and the Coriolis force, can significantly improve the forecasts. In addition, data assimilation is required to improve the simulations and correct the dissipative effect of the numerical scheme.

The sediment transport modelling is of major interest in terms of applications, in particular to estimate the sustainability of facilities with silt or scour, such as canals and bridges. Dredging or filling-up operations are expensive and generally not efficient in the long term. The objective is to determine a configuration almost stable for the facilities. In addition, it is also important to determine the impact of major events like emptying dam which is aimed at evacuating the sediments in the dam reservoir and requires a large discharge. However, the downstream impact should be measured in terms of turbidity, river morphology and flood.

## 4.3. Hydrological disasters

It is a violent, sudden and destructive flow. Between 1996 and 2005, nearly 80% of natural disasters in the world have meteorological or hydrological origins. The main interest of their study is to predict the areas in which they may occur most probably and to prevent damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worst damages. For example, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last require the use of reduced models like the shallow water equations. In urban areas, the flow can be largely impacted by the debris, in particular cars, and this requires fluid/structure interactions be well understood. Moreover, underground flows, in particular in sewers, can accelerate and amplify the flow. To take them into account, the model and the numerical resolution should be able to treat the transition between free surface and underground flows.

Tsunamis are another hydrological disaster largely studied. Even if the propagation of the wave is globally well described by the shallow water model in oceans, it is no longer the case close to the epicenter and in the coastal zone where the bathymetry leads to vertical accretions and produces substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

While viscous effects can often be neglected in water flows, they have to be taken into account in situations such as avalanches, debris flows, pyroclastic flows, erosion processes, ...*i.e.* when the fluid rheology becomes more complex. Gravity driven granular flows consist of solid particles commonly mixed with an interstitial lighter fluid (liquid or gas) that may interact with the grains and decrease the intensity of their contacts, thus reducing energy dissipation and favoring propagation. Examples include subaerial or subaqueous rock avalanches (e.g. landslides).

#### **4.4. Biodiversity and culture**

Nowadays, simulations of the hydrodynamic regime of a river, a lake or an estuary, are not restricted to the determination of the water depth and the fluid velocity. They have to predict the distribution and evolution of external quantities such as pollutants, biological species or sediment concentration.

The potential of micro-algae as a source of biofuel and as a technological solution for CO<sub>2</sub> fixation is the subject of intense academic and industrial research. Large-scale production of micro-algae has potential for biofuel applications owing to the high productivity that can be attained in high-rate raceway ponds. One of the key challenges in the production of micro-algae is to maximize algae growth with respect to the exogenous energy that must be used (paddlewheel, pumps, ...). There is a large number of parameters that need to be optimized (characteristics of the biological species, raceway shape, stirring provided by the paddlewheel). Consequently our strategy is to develop efficient models and numerical tools to reproduce the flow induced by the paddlewheel and the evolution of the biological species within this flow. Here, mathematical models can greatly help us reduce experimental costs. Owing to the high heterogeneity of raceways due to gradients of temperature, light intensity and nutrient availability through water height, we cannot use depth-averaged models. We adopt instead more accurate multilayer models that have recently been proposed. However, it is clear that many complex physical phenomena have to be added to our model, such as the effect of sunlight on water temperature and density, evaporation and external forcing.

Many problems previously mentioned also arise in larger scale systems like lakes. Hydrodynamics of lakes is mainly governed by geophysical forcing terms: wind, temperature variations, ...

#### **4.5. Sustainable energy**

One of the booming lines of business is the field of renewable and decarbonated energies. In particular in the marine realm, several processes have been proposed in order to produce electricity thanks to the recovering of wave, tidal and current energies. We may mention water-turbines, buoys turning variations of the water height into electricity or turbines motioned by currents. Although these processes produce an amount of energy which is less substantial than in thermal or nuclear power plants, they have smaller dimensions and can be set up more easily.

The fluid energy has kinetic and potential parts. The buoys use the potential energy whereas the water-turbines are activated by currents. To become economically relevant, these systems need to be optimized in order to improve their productivity. While for the construction of a harbour, the goal is to minimize swell, in our framework we intend to maximize the wave energy.

This is a complex and original issue which requires a fine model of energy exchanges and efficient numerical tools. In a second step, the optimisation of parameters that can be changed in real-life, such as bottom bathymetry and buoy shape, must be studied. Eventually, physical experiments will be necessary for the validation.

#### **4.6. Urban environment**

The urban environment is essentially studied for air and noise pollutions. Air pollution levels and noise pollution levels vary a lot from one street to next. The simulations are therefore carried out at street resolution and take into account the city geometry. The associated numerical models are subject to large uncertainties. Their input parameters, e.g. pollution emissions from road traffic, are also uncertain. Quantifying

the simulation uncertainties is challenging because of the high computational costs of the numerical models. An appealing approach in this context is the use of metamodels, from which ensembles of simulations can be generated for uncertainty quantification.

The simulation uncertainties can be reduced by the assimilation of fixed and mobile sensors. High-quality fixed monitoring sensors are deployed in cities, and an increasing number of mobile sensors are added to the observational networks. Even smartphones can be used as noise sensors and dramatically increase the spatial coverage of the observations. The processing and assimilation of the observations raises many questions regarding the quality of the measurements and the design of the network of sensors.

## **4.7. SmartCity**

There is a growing interest for environmental problems at city scale, where a large part of the population is concentrated and where major pollutions can occur. Numerical simulation is well established to study the urban environment, *e.g.* for road traffic modelling. As part of the smartcity movement, an increasing number of sensors collect measurements, at traditional fixed observation stations, but also on mobile devices, like smartphones. They must properly be taken into account given their number but also their potential low quality.

Practical applications include air pollution and noise pollution. These directly relate to road traffic. Data assimilation and uncertainty propagation are key topics in these applications.

## **CASTOR Project-Team**

### **4. Application Domains**

#### **4.1. Nuclear fusion**

The activity of Castor is mainly applied to nuclear fusion, in particular on the WEST, JET and ITER Tokamaks. Several tools developed in the project are used on those machines like equilibrium reconstruction, ELMs simulations...

## **COFFEE Project-Team**

### **4. Application Domains**

#### **4.1. Multiphase flows in porous media**

Our research focuses on the numerical modeling of multiphase porous media flows accounting for complex geology and for nonlinear and multi-physics couplings. It is applied to various problems in the field of energy such as the simulation of geothermal systems in collaboration with BRGM, of nuclear waste repositories in collaboration with Andra, and of oil and gas recovery in collaboration with Total. Our research directions include the development of advanced numerical schemes adapted to polyhedral meshes and highly heterogeneous media in order to represent more accurately complex geologies. A special focus is made on the modeling of multiphase flows in network of faults or fractures represented as interfaces of co-dimension one coupled to the surrounding matrix. We also investigate nonlinear solvers adapted to the nonlinear couplings between gravity, capillary and viscous forces in highly heterogeneous porous media. In the same line, we study new domain decomposition algorithms to couple non-isothermal compositional liquid gas flows in a porous medium with free gas flows occurring at the interface between the ventilation gallery and the nuclear waste repository or between a geothermal reservoir and the atmosphere.

#### **4.2. Particulate and mixture flows**

We investigate fluid mechanics models referred to as “multi-fluids” flows. A large part of our activity is more specifically concerned with the case where a disperse phase interacts with a dense phase. Such flows arise in numerous applications, like for pollutant transport and dispersion, the combustion of fuel particles in air, the modelling of fluidized beds, the dynamic of sprays and in particular biosprays with medical applications, engine fine particles emission... There are many possible modelings of such flows: microscopic models where the two phases occupy distinct domains and where the coupling arises through intricate interface conditions; macroscopic models which are of hydrodynamic (multiphase) type, involving non standard state laws, possibly with non conservative terms, and the so-called mesoscopic models. The latter are based on Eulerian-Lagrangian description where the disperse phase is described by a particle distribution function in phase space. Following this path we are led to a Vlasov-like equation coupled to a system describing the evolution of the dense phase that is either the Euler or the Navier-Stokes equations. It turns out that the leading effect in such models is the drag force. However, the role of other terms, of more or less phenomenological nature, deserves to be discussed (close packing terms, lift term, Basset force...). Of course the fluid/kinetic model is interesting in itself and needs further analysis and dedicated numerical schemes. In particular, in collaboration with the Atomic Energy Commission (CEA), we have proposed a semi-Lagrangian scheme for the simulation of particulate flows, extending the framework established in plasma physics to such flows.

We also think it is worthwhile to identify hydrodynamic regimes: it leads to discuss hierarchies of coupled hydrodynamic systems, the nature of which could be quite intriguing and original, while they share some common features of the porous media problems. We are particularly interested in revisiting the modeling of mixture flows through the viewpoint of kinetic models and hydrodynamic regimes. We propose to revisit the derivation of new mixture models, generalizing Kazhikov-Smagulov equations, through hydrodynamic asymptotics. The model is of “hybrid” type in the sense that the constraint reduces to the standard incompressibility condition when the disperse phase is absent, while it involves derivatives of the particle volume fraction when the disperse phase is present.

### **4.3. Biological degradation, biofilms formation and algae proliferation**

Members of the team have started an original research program devoted to biofilms formation and algae proliferation. We started working on this subject through a collaboration with Roberto Natalini and a group of experts in Firenze interested in preventing damages on historical monuments. It is also motivated by *Ostreopsis* proliferation in the Mediterranean Sea. The multidisciplinary character of this research relies on discussions with researchers of the Oceanography Laboratory in Villefranche-sur-Mer, a leading marine research unit, and the Inria team BIOCORE, led by J-L Gouzé. This research was supported by a ANR-project, led by M. Ribot, and it was the main topic of the PhD thesis of B. Polizzi. It continues through a collaboration with INRA (Jouy-en-Josas), dealing with the modeling of flows in the gut and with the aim of describing the formation and stability of mucus layers.

**FLUMINANCE Project-Team (section vide)**

## LEMON Team

# 4. Application Domains

## 4.1. Simulation of extreme events

The models and methods developed within this research line aim to support decision-making in the field of flood crisis management. This concerns various types of floods.

- rainfall-induced floods, stemming from intense rainfall events (see 3.2.3 , 3.4.1 , 3.4.2 ),
- fluvial floods
- fast rising flood waves induced by dam/dike break or tsunami waves (see 3.3.1 , 3.2.2 ),
- coastal submersion (storm surges).

The aforementioned models and methods can be used at the three stages of the flood crisis.

- Before the crisis, models can be run offline to assess the vulnerability and resilience of alternative urbanization schemes, rescue actions and mitigation policies to various meteorological/coastal scenarios. This includes the simulation of extreme forcings (wind, wave and/or rainfall fields) from a limited set of available records, see subsection 3.3.2 .
- During the crisis, either fast-running porosity models or pre-computed scenarios may be used to support real-time decision-making for rescue actions.
- After the crisis, data and experience collected at the crisis stage may be used to enrich the simulation database, better parametrize the models and methods.

## 4.2. Marine and coastal systems

LEMON will consider interactions between various processes (and the corresponding models) in coastal oceanography. Our guideline is the design and implementation of accurate numerical methods to simulate physical and/or ecological processes involving various time and space scales:

- modelling of waves in several regimes of the coastal area (shoaling, breaking, running, etc.) implies the implementation of numerical models involving various physical regimes (2D, 3D, hydrostatic, dispersive, etc.) which cannot be studied independently (in particular in the framework of marine energies). Each of the corresponding models has its own specificity, requirements and experts. Hence the difficulty to build a multiphysics (coupled) global ocean system at the coastal scale.
- for most coastal marine organisms, including algae, invertebrates and fish, dispersal, genetic connectivity and the replenishment of coastal populations depend on the character of the coastal ocean. Modelling this dispersal process and larval transport across the surfzone is at the foremost of marine conservation ecology and requires advanced ocean modelling (in addition to the specific biological processes).

LEMON has a strong expertise on mathematical modelling and coupling techniques (borrowed from domain decomposition models) that will be intensely used in these frameworks. As far as extreme events are concerned, we want to build a data generator which will allow the study of the effects of various scenarios on impact measures. To do so, we will first analyze data simulated from a physical system and detect extreme events. The latter will be used to generate and calibrate as many realistic extreme events as necessary.



## **MAGIQUE-3D Project-Team**

### **4. Application Domains**

#### **4.1. Seismic Imaging**

The main objective of modern seismic processing is to find the best representation of the subsurface that can fit the data recorded during the seismic acquisition survey. In this context, the seismic wave equation is the most appropriate mathematical model. Numerous research programs and related publications have been devoted to this equation. An acoustic representation is suitable if the waves propagate in a fluid. But the subsurface does not contain fluids only and the acoustic representation is not sufficient in the general case. Indeed the acoustic wave equation does not take some waves into account, for instance shear waves, turning waves or the multiples that are generated after several reflections at the interfaces between the different layers of the geological model. It is then necessary to consider a mathematical model that is more complex and resolution techniques that can model such waves. The elastic or viscoelastic wave equations are then reference models, but they are much more difficult to solve, in particular in the 3D case. Hence, we need to develop new high-performance approximation methods.

Reflection seismics is an indirect measurement technique that consists in recording echoes produced by the propagation of a seismic wave in a geological model. This wave is created artificially during seismic acquisition surveys. These echoes (i.e., reflections) are generated by the heterogeneities of the model. For instance, if the seismic wave propagates from a clay layer to sand, one will observe a sharp reflected signal in the seismic data recorded in the field. One then talks about reflection seismics if the wave is reflected at the interface between the two media, or talks about seismic refraction if the wave is transmitted along the interface. The arrival time of the echo enables one to locate the position of this transition, and the amplitude of the echo gives information on some physical parameters of the two geological media that are in contact. The first petroleum exploration surveys were performed at the beginning of the 1920's and for instance, the Orchard Salt Dome in Texas (USA) was discovered in 1924 by the seismic-reflection method.

#### **4.2. Imaging complex media with ultrasonic waves**

The acoustic behavior of heterogeneous or composite materials attracts considerable excitement. Indeed, their acoustic response may be extremely different from the single constituents responses. In particular, dispersions of resonators in a matrix are the object of large research efforts, both experimentally and theoretically. However it is still a challenge to dispose of numerical tools with sufficient abilities to deal with the simulation and imaging of such materials behavior. Indeed, not only acoustic simulations are very time-consuming, but they have to be performed on realistic enough solution domains, i.e. domains which capture well enough the structural features of the considered materials.

This collaboration with I2M, University of Bordeaux aims at addressing this type of challenges by developing numerical and experimental tools in order to understand the propagation of ultrasonic waves in complex media, image these media, and in the future, help design composite materials for industrial purposes.

#### **4.3. Helioseismology**

This collaboration with the Max Planck Institute for Solar System, Göttingen, Germany, which started in 2014, aims at designing efficient numerical methods for the wave propagation problems that arise in helioseismology in the context of inverse problems. The final goal is to retrieve information about the structure of the Sun i.e. inner properties such as density or pressure via the inversion of a wave propagation problem. Acoustic waves propagate inside the Sun which, in a first approximation and regarding the time scales of physical phenomena, can be considered as a moving fluid medium with constant velocity of motion. Some other simplifications lead to computational saving, such as supposing a radial or axisymmetric geometry of the Sun.

Aeroacoustic equations must be adapted and efficiently solved in this context, this has been done in the finite elements code Montjoie. In other situations, a full 3D simulation is required and demands large computational resources. Ultimately, we aim at modeling the coupling with gravity potential and electromagnetic waves (MHD equations) in order to be able to better understand Sun spots.

## **SERENA Project-Team**

### **4. Application Domains**

#### **4.1. Multiphase flows and transport of contaminants in the subsurface**

- subsurface depollution after chemical leakage
- nuclear waste disposal in deep underground repositories
- flow in large scale discrete fracture networks
- production of oil and gas

#### **4.2. Industrial risks in energy production**

- Stokes and Navier–Stokes flows related to nuclear reactor operation
- reduced-order models for valves related to nuclear reactor operation
- plasticity and large deformations for mechanical components related to nuclear reactor operation
- seismic wave propagation for detection and protection
- electromagnetism for interfaces between dielectrics and negative metamaterials

#### **4.3. Computational quantum chemistry**

- guaranteed bounds for ground-state energy (eigenvalues) and ground-state density matrix (eigenvectors) in first-principle molecular simulation
- application to Laplace, Gross–Pitaevskii, Kohn–Sham, and Schrödinger models

**STEPP Project-Team (section vide)**

## TONUS Team

# 4. Application Domains

## 4.1. Controlled fusion and ITER

The search for alternative energy sources is a major issue for the future. Among others, controlled thermonuclear fusion in a hot hydrogen plasma is a promising possibility. The principle is to confine the plasma in a toroidal chamber, called a tokamak, and to attain the necessary temperatures to sustain nuclear fusion reactions. The International Thermonuclear Experimental Reactor (ITER) is a tokamak being constructed in Cadarache, France. This was the result of a joint decision by an international consortium made of the European Union, Canada, USA, Japan, Russia, South Korea, India and China. ITER is a huge project. As of today, the budget is estimated at 20 billion euros. The first plasma shot is planned for 2020 and the first deuterium-tritium operation for 2027. Many technical and conceptual difficulties have to be overcome before the actual exploitation of fusion energy. Consequently, much research has been carried out around magnetically confined fusion. Among these studies, it is important to carry out computer simulations of the burning plasma. Thus, mathematicians and computer scientists are also needed in the design of ITER. The reliability and the precision of numerical simulations allow a better understanding of the physical phenomena and thus would lead to better designs. TONUS's main involvement is in such research. The required temperatures to attain fusion are very high, of the order of a hundred million degrees. Thus it is imperative to prevent the plasma from touching the tokamak inner walls. This confinement is obtained thanks to intense magnetic fields. The magnetic field is created by poloidal coils, which generate the toroidal component of the field. The toroidal plasma current also induces a poloidal component of the magnetic field that twists the magnetic field lines. The twisting is very important for the stability of the plasma. The idea goes back to research by Tamm and Sakharov, two Russian physicists, in the 50's. Other devices are essential for the proper operation of the tokamak: divertor for collecting the escaping particles, microwave heating for reaching higher temperatures, fuel injector for sustaining the fusion reactions, toroidal coils for controlling instabilities, etc.

## 4.2. Other applications

The software and numerical methods that we develop can also be applied to other fields of physics or of engineering.

- For instance, we have a collaboration with the company AxesSim in Strasbourg for the development of efficient Discontinuous Galerkin (DG) solvers on hybrid computers. The applications are electromagnetic simulations for the conception of antennas, electronic devices or aircraft electromagnetic compatibility.
- The acoustic conception of large rooms requires huge numerical simulations. It is not always possible to solve the full wave equation and many reduced acoustic models have been developed. A popular model consists in considering "acoustic" particles moving at the speed of sound. The resulting Partial Differential Equation (PDE) is very similar to the Vlasov equation. The same modelling is used in radiation theory. We have started to work on the reduction of the acoustic particles model and realized that our reduction approach perfectly applies to this situation. A PhD with CEREMA (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) has started in October 2015 (PhD of Pierre Gerhard). The objective is to investigate the model reduction and to implement the resulting acoustic model in our DG solver.
- In September 2017, we started a collaboration with EDF Chatou (PhD of Lucie Quibel) on the modelling of multiphase fluids with complex equations of state. The goal is to simulate the high temperature liquid-vapor flow occurring in a nuclear plant. Among others, we will apply our recent kinetic method for designing efficient implicit schemes for this kind of flows.

## **BIOCORE Project-Team**

# **4. Application Domains**

## **4.1. Bioenergy**

Finding sources of renewable energy is a key challenge for our society. We contribute to this topic through two main domains for which a strong and acknowledged expertise has been acquired over the years. First, we consider anaerobic digesters, the field of expertise of the members of the team at the Laboratory of Environmental Biotechnology (LBE), for the production of methane and/or biohydrogen from organic wastes. The main difficulty is to make these processes more reliable and exploit more efficiently the produced biogas by regulating both its quality and quantity despite high variability in the influent wastes. One of the specific applications that needs to be tackled is the production of biogas in a plant when the incoming organic waste results from the mixing of a finite number of substrates. The development of control laws that optimize the input mix of the substrates as a function of the actual state of the system is a key challenge for the viability of this industry.

The second topic consists in growing microalgae, the field of expertise of the members of the team at the Oceanographic Laboratory of Villefranche-sur-Mer (LOV), to produce biofuel. These microorganisms can synthesize lipids with a much higher productivity than terrestrial oleaginous species. The difficulty is to better understand the involved processes, which are mainly transient, to stimulate and optimize them on the basis of modeling and control strategies. Predicting and optimizing the productivity reached by these promising systems in conditions where light received by each cell is strongly related to hydrodynamics, is a crucial challenge.

Finally, for the energy balance of the process, it is important to couple microalgae and anaerobic digestion to optimize the solar energy that can be recovered from microalgae, as was explored within the ANR Symbiose project (2009-2012) [3].

## **4.2. CO<sub>2</sub> fixation and fluxes**

Phytoplanktonic species, which assimilate CO<sub>2</sub> during photosynthesis, have received a lot of attention in the last years. Microalgal based processes have been developed in order to mitigate industrial CO<sub>2</sub>. As for biofuel productions, many problems arise when dealing with microalgae which are more complex than bacteria or yeasts. Several models have been developed within our team to predict the CO<sub>2</sub> uptake in conditions of variable light and nitrogen availability. The first modeling challenge in that context consists in taking temperature effects and light gradient into account.

The second challenge consists in exploiting the microalgal bioreactors which have been developed in the framework of the quantification of carbon fluxes between ocean and atmospheres. The SEMPO platform (simulator of variable environment computer controlled), developed within the LOV team, has been designed to reproduce natural conditions that can take place in the sea and to accurately measure the cells behavior. This platform, for which our team has developed models and control methods over the years, is an original and unique tool to develop relevant models which stay valid in dynamic conditions. It is worth noting that a better knowledge of the photosynthetic mechanisms and improved photosynthesis models will benefit both thematics: CO<sub>2</sub> mitigation and carbon fluxes predictions in the sea.

## **4.3. Biological control for plants and micro-plants production systems**

This research concentrates on the protection of cultures of photosynthetic organisms against their pests or their competitors. The cultures we study are crop and micro-algae productions. In both cases, the devices are more or less open to the outside, depending on the application (greenhouse/field, photobioreactor/raceway), so that they may give access to harmful pathogens and invading species. We opt for protecting the culture through the use of biocontrol in a broad sense.

In crop production, biocontrol is indeed a very promising alternative to reduce pesticide use: it helps protecting the environment, as well as the health of consumers and producers; it limits the development of resistance (in comparison to chemicals). The use of biocontrol agents, which are, generically, natural enemies (predators, parasitoids or pathogens) of crop pests [84], is however not widespread yet because it often lacks efficiency in real-life crop production systems (while its efficiency in the laboratory is much higher) and can fail to be economically competitive. Resistant crops are also used instead of pesticides to control pests and pathogens, but the latter eventually more or less rapidly overcome the resistance, so these crops need to be replaced by new resistant crops. As resistant genes are a potentially limited resource, a challenge is to ensure the durability of crop resistance. Our objective is to propose models that would help to explain which factors are locks that prevent the smooth transition from the laboratory to the agricultural crop, as well as develop new methods for the optimal deployment of the pests natural enemies and of crop resistance.

Microalgae production is faced with exactly the same problems since predators of the produced microalgae (e.g. zooplankton) or simply other species of microalgae can invade the photobioreactors and outcompete or eradicate the one that we wish to produce. Methods need therefore to be proposed for fighting the invading species; this could be done by introducing predators of the pest and so keeping it under control, or by controlling the conditions of culture in order to reduce the possibility of invasion; the design of such methods could greatly take advantage of our knowledge developed in crop protection since the problems and models are related.

#### **4.4. Biological depollution**

These works will be carried out with the LBE , mainly on anaerobic treatment plants. This process, despite its strong advantages (methane production and reduced sludge production) can have several locally stable equilibria. In this sense, proposing reliable strategies to stabilize and optimise this process is a key issue. Because of the recent (re)development of anaerobic digestion, it is crucial to propose validated supervision algorithms for this technology. A problem of growing importance is to take benefit of various waste sources in order to adapt the substrate quality to the bacterial biomass activity and finally optimize the process. This generates new research topics for designing strategies to manage the fluxes of the various substrate sources meeting at the same time the depollution norms and providing a biogas of constant quality. In the past years, we have developed models of increasing complexity. However there is a key step that must be considered in the future: how to integrate the knowledge of the metabolisms in such models which represent the evolution of several hundreds bacterial species? How to improve the models integrating this two dimensional levels of complexity? With this perspective, we wish to better represent the competition between the bacterial species, and drive this competition in order to maintain, in the process, the species with the highest depollution capability. This approach, initiated in [88] must be extended from a theoretical point of view and validated experimentally.

## CARMEN Project-Team

### 4. Application Domains

#### 4.1. Scientific context: the LIRYC

The University Hospital of Bordeaux (*CHU de Bordeaux*) is equipped with a specialized cardiology hospital, the *Hôpital Cardiologique du Haut-Lévêque*, where the group of Professor Michel Haïssaguerre has established itself as a global leader in the field of cardiac electrophysiology [58], [57], [47]. Their discoveries in the area of atrial fibrillation and sudden cardiac death syndromes are widely acclaimed, and the group is a national and international referral center for treatment of cardiac arrhythmia. Thus the group also sees large numbers of patients with rare cardiac diseases.

In 2011 the group has won the competition for a 40 million euro *Investissements d'Avenir* grant for the establishment of IHU Liryc, an institute that combines clinical, experimental, and numerical research in the area of cardiac arrhythmia (<http://ihu-liryc.fr>). The institute works in all areas of modern cardiac electrophysiology: atrial arrhythmias, sudden death due to ventricular fibrillation, heart failure related to ventricular dyssynchrony, and metabolic disorders. It is recognized as one of the most important centers worldwide in this area.

The Carmen team was founded to partner with IHU Liryc. We bring applied mathematics and scientific computing closer to experimental and clinical cardiac electrophysiology. In collaboration with experimental and clinical researchers at Liry we work to enhance fundamental knowledge of the normal and abnormal cardiac electrical activity and of the patterns of the electrocardiogram, and we develop new simulation tools for training, biological, and clinical applications.

#### 4.2. Basic experimental electrophysiology

Our modeling is carried out in coordination with the experimental teams from IHU Liryc. It help to write new concepts concerning the multiscale organisation of the cardiac action potentials that will serve our understanding in many electrical pathologies. For example, we model the structural heterogeneities at the cellular scale [31], and at an intermediate scale between the cellular and tissue scales.

At the atrial level, we apply our models to understand the mechanisms of complex arrhythmias and the relation with the heterogeneities at the insertion of the pulmonary veins. We will model the heterogeneities specific to the atria, like fibrosis or fatty infiltration [34] [56]. These heterogeneities ara thought to play a major role in the development of atrial fibrillation.

At the ventricular level, we focus on (1) modeling the complex coupling between the Purkinje network and the ventricles, which is supposed to play a major role in sudden cardiac death, and (2) modeling the heteogeneities related to the complex organization and disorganization of the myocytes and fibroblasts, which is important in the study of infarct scars for instance.

#### 4.3. Clinical electrophysiology

Treatment of cardiac arrhythmia is possible by pharmacological means, by implantation of pacemakers and defibrillators, and by curative ablation of diseased tissue by local heating or freezing. In particular the ablative therapies create challenges that can be addressed by numerical means. Cardiologists would like to know, preferably by noninvasive means, where an arrhythmia originates and by what mechanism it is sustained.

We address this issue in the first place using inverse models, which attempt to estimate the cardiac activity from a (high-density) electrocardiogram. A new project aims at performing this estimation on-site in the catheterization laboratory and presenting the results, together with the cardiac anatomy, on the screen that the cardiologist uses to monitor the catheter positions [40].

An important prerequisite for this kind of interventions and for inverse modeling is the creation of anatomical models from imaging data. The Carmen team contributes to better and more efficient segmentation and meshing through the IDAM project.



**DRACULA Project-Team (section vide)**

## **M3DISIM Project-Team**

# **4. Application Domains**

## **4.1. Clinical applications**

After several validation steps – based on clinical and experimental data – we have reached the point of having validated the heart model in a pre-clinical context where we have combined direct and inverse modeling in order to bring predictive answers on specific patient states. For example, we have demonstrated the predictive ability of our model to set up pacemaker devices for a specific patient in cardiac resynchronization therapies, see [11]. We have also used our parametric estimation procedure to provide a quantitative characterization of an infarct in a clinical experiment performed with pigs, see [3].

## MAMBA Project-Team

# 4. Application Domains

## 4.1. Introduction

The team has two main application-driven research axes.

*Applicative axis 1* focuses on cancer, an application on which almost all team members work, with various approaches. A main focus of the team is to study cancer as a Darwinian evolutionary phenomenon in phenotype-structured cell populations. Optimal control methods take into account the two main pitfalls of clinical cancer therapeutics, namely unwanted toxic side effects in healthy cell populations and drug resistance in cancer cell populations. Other studies concern telomere shortening, and multi-scale models.

*Applicative axis 2* is devoted to growth, evolution and regeneration in populations and tissues. It involves protein aggregation and fragmentation models for neurodegenerative diseases (prion, Alzheimer), organ modeling, mainly of the liver, its damages induced by toxic molecules, and its regeneration after toxic insult. Newcomers in this applicative field are epidemiological modeling of propagation of insect vector-borne diseases by reaction-diffusion equations and of their optimal control, bacterial growth and wound healing.

## 4.2. Applicative axis 1: Focus on cancer

### Personnel

Luis Almeida, Cécile Carrère, Jean Clairambault, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard.

### Project-team positioning

The MAMBA team designs and analyses mathematical models of tumor growth and therapy, at the cell population level, using agent-based or partial differential equations, with special interest in methodologies for therapeutic optimisation using combined anticancer drug treatments. Rather than, or not only, modeling the effect of drugs on molecular targets, we represent these effects by their *functional* consequences on the fate of healthy and cancer cell populations: proliferation (velocity of the cell division cycle, decreasing it, e.g., by antagonizing growth factor receptors), apoptosis, cell death or senescence.

Our goal in doing this is to circumvent the two main issues of anticancer therapy in the clinic, namely unwanted toxic side effects in populations of healthy cells and emergence of drug-induced drug resistance in cancer cell populations. This point of view leads us to take into account phenomena of transient and reversible resistance, observed in many cancer cell populations, by designing and analyzing models of cell populations structured in continuous phenotypes, relevant for the description of the behavior of cell populations exposed to drugs: either degree of resistance to a given drug, or potential of resistance to drug-induced stress, proliferation potential, and plasticity.

Such modeling options naturally lead us to take into account in a continuous way (i.e., by continuous-valued phenotype or relevant gene expression) the wide phenotypic heterogeneity of cancer cell populations. They also lead us to adopt the point of view of *adaptive dynamics* according to which characteristic traits of cell populations evolve with tumor environmental pressure (drugs, cytokines or metabolic conditions, mechanical stress and spatial conditions), in particular from drug sensitivity to resistance. This position is original on the international scene of teams dealing with drug resistance in cancer.

### Scientific achievements

**Modeling Acute Myeloid Leukemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations**

In collaboration with Catherine Bonnet (Inria DISCO, Saclay) and François Delhommeau (St Antoine hospital in Paris), together with DISCO PhD students José Luis Avila Alonso and Walid Djema, this theme has led to common published proceedings of conferences: IFAC, ACC, CDC, MTNS [66], [67], [68], [77], [91], [65]. These works study the stability of the haematopoietic system and its possible restabilization by combinations of anticancer drugs with functional targets on cell populations: proliferation, apoptosis, differentiation.

#### **Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control**

We tackle the problem to represent and inhibit - using optimal control algorithms, in collaboration with Emmanuel Trélat, proposed Inria team CAGE - drug-induced drug resistance in cancer cell populations. This theme, presently at the core of our works on cancer modeling with a evolutionary perspective on tumor heterogeneity, is documented in a series of articles [88], [89], [115], [116], [118]. Taking into account the two main pitfalls of cancer therapy, unwanted side effects on healthy cells and evolution towards resistance in cancer cells, it has attracted to our team the interest of several teams of biologists, with whom we have undertaken common collaborative works, funded by laureate answers to national calls (see ITMO Cancer HTE call).

This theme is also at the origin of methodological developments (see Research axis 1). In collaboration with Shensi Shen from Institut Gustave Roussy and François Vallette from Université de Nantes, we aim to develop simple non-spatial models to understand the mechanisms of drug resistance acquisition -and lost- in melanoma and glioblastoma. The models are systematically compared with in vitro and in vivo data generated by our collaborators and treated via image processing techniques developed in the team.

#### **Senescence modeling by telomere shortening**

In many animals, aging tissues accumulate senescent cells, a process which is beneficial to protect from cancer in the young organism. In collaboration with Teresa Teixeira and Zhou Xu from IBCP, we proposed a mathematical model based on the molecular mechanisms of telomere replication and shortening and fitted it on individual lineages of senescent *Saccharomyces cerevisiae* cells, in order to decipher the causes of heterogeneity in replicative senescence [79].

#### **Biomechanically mediated growth control of cancer cells**

Model simulations indicate that the response of growing cell populations on mechanical stress follows a simple universal functional relationship and is predictable over different cell lines and growth conditions despite the response curves look largely different. We developed a hybrid model strategy in which cells were represented by coarse-grained individual units calibrated in a high resolution cell model and parameterized each model cell by measurable biophysical and cell-biological parameters. Cell cycle progression in our model is controlled by volumetric strain, the latter being derived from a bio-mechanical relation between applied pressure and cell compressibility. After parameter calibration from experiments with mouse colon carcinoma cells growing against the resistance of an elastic alginate capsule, the model adequately predicts the growth curve in i) soft and rigid capsules, ii) in different experimental conditions where the mechanical stress is generated by osmosis via a high molecular weight dextran solution, and iii) for other cell types with different growth kinetics. Our model simulation results suggest that the growth response of cell population upon externally applied mechanical stress is the same, as it can be quantitatively predicted using the same growth progression function [44].

#### **Collaborations**

- AML modelling: **Catherine Bonnet**, DISCO Inria team, Saclay, and **François Delhommeau**, INSERM St Antoine (also collaborator in the INSERM HTE laureate project EcoAML, see below).
- INSERM HTE laureate project MoGIImaging, headed by E. Moyal (Toulouse): **François Vallette**, CRCNA and INSERM Nantes
- INSERM HTE laureate project EcoAML, headed by **François Delhommeau**, INSERM St Antoine: François Delhommeau, Thierry Jaffredo (IBPS), Delphine Salort (LCQB-IBPS)
- Adaptive dynamics to model drug resistance and optimal control to circumvent it:

**Alexandre Escargueil, Michèle Sabbah** (1 PhD thesis in common), St Antoine Hospital, Paris

**Emmanuel Trélat** (1 PhD thesis in common) at Inria team CAGE and Laboratoire Jacques-Louis Lions at Sorbonne Université.

**Frédéric Thomas** at CREEC, Montpellier.

**Tommaso Lorenzi** (Univ. of St Andrews).

- Telomere shortening: **Teresa Teixeira** and **Zhou Xu** (IBCP, Paris), **Philippe Robert** (Inria RAP).
- TRAIL treatment: **Gregory Batt**, Inria Saclay and Inst. Pasteur (France)
- Biomechanical control of cancer cells: **Pierre Nassoy**, Bioimaging and Optofluidics Group, LP2N – UMR 5298, IOGS, CNRS & University of Bordeaux

### 4.3. Applicative axis 2: Growth, evolution and regeneration in populations and tissues

#### Personnel

Luis Almeida, Pierre-Alexandre Bliman, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil, Philippe Robert

#### Project-team positioning

The applications in this category span very different subjects from amyloid diseases, dengue fever, wound healing, liver regeneration and toxicity, up to bacterial growth and development of organisms. As the applications, the methods span a wide range. Those concerning identification of models and parameters with regard to data have partially been outlined in axis 3. Focus in this axis is on the model contribution to the biologically and/or medically relevant insights and aspects.

Liver-related modeling is partially performed within the Inria team MIMESIS (Strasbourg) with the focus on real-time, patient-specific biomechanical liver models to guide surgery and surgeons. Internationally, spatial temporal liver related models are developed in Fraunhofer MEVIS (Bremen), by T. Ricken (University of Stuttgart), and P. Segers group (Leuven). Different from these, Mamba has a strong focus on spatial-temporal modeling on the histological scale, integration of molecular processes in each individual cell, and single-cell (agent) based models [32], [30], [143]. Works by Schliess [139], [101] have been highlighted in editorials.

Mathematical modeling of protein aggregation is a relatively recent domain, only a few other groups have emerged yet; among them we can cite the Inria team Dracula, with whom we are in close contact, and e.g., the work by Jean-Michel Coron (Sorbonne Université) and Monique Chyba (Hawaii, USA) in control, and Suzanne Sindi (USA) for the modeling of the yeast prion. We have interactions with all these groups and organized a workshop in June 2017, gathering both the biophysics and applied mathematics communities.

#### Scientific achievements

##### Amyloid disease

Application to protein aggregation in amyloid diseases is a long-standing interest of Mamba, dating back to 2010 [85], and developed through the collaboration with Human Rezaei's team at Inra. More recently, with Wei-Feng Xue in Canterbury, we investigated the intrinsic variability among identical experiments of nucleation [93], [100], Sarah Eugène's Ph.D subject (co-supervised by Philippe Robert) [99].

In collaboration with Tom Banks first [69], [70] and then Philippe Moireau, we developed quantitative comparisons between model and data. Through data assimilation and statistical methods [63], we proposed new models and mechanisms.

##### Biological control of arboviruses

Sterile Insect Technique (SIT) [98] is a biological control method relying on massive releases of sterile male insects into the wild. The latter compete with wild males to mate with the females, and induce no offspring to the latter, thus reducing the next generation's population. This can result in a progressive reduction, or even disparition, of the target population.

A related technique is based on the infection by *Wolbachia* [104]. This symbiotic bacterium is maternally transmitted from infected females to their offspring, but induces *cytoplasmic incompatibility* [141], [80]: mating between infected males and uninfected females gives no offspring. Releases of *Wolbachia* infected males alone is thus comparable to classical SIT.

On the other hand, releasing both infected males and females in sufficient quantity may result in infection of the wild population. This gives rise to an interesting new control principle, as *Wolbachia* has been shown to severely reduce the insect vectorial ability to transmit dengue, zika or chikungunya, indirectly by lifespan and fertility reduction, and directly by reducing the ability of the viruses to proliferate within the organism [121].

We proposed new insights on the practical and theoretical issues raised by the implementation of the previous methods.

### **Wound healing 1: epithelial tissues**

We studied cell motion in epithelial gap closure, a form of collective cell migration that is a very widespread phenomenon both during development and adult life - it is essential for both the formation and for the maintenance of epithelial layers. Due to their importance, *in vivo* wound healing and morphogenetic movements involving closure of holes in epithelia have been the object of many studies. In our works <sup>0</sup> we considered wound healing and epithelial gap closure in both *in vivo* (in particular *Drosophila* pupa) and *in vitro* (MDCK cell and human keratinocytes). We found some similarities in the geometry dependence of the wound closure strategies between these two situations, indicating the existence of conserved mechanisms that should be widespread across living beings. We are concentrating on the study of actin cable formation.

### **Wound healing 2: adipose tissues**

After injury, if regeneration can be observed in hydra, planaria and some vertebrates, regeneration is rare in mammals and particularly in humans. In this research axis, we investigated the mechanisms by which biological tissues recover after injury. We explored this question on adipose tissue, using the mathematical framework recently developed in [134]. Our assumption is that simple mechanical cues between the Extra-Cellular Matrix (ECM) and differentiated cells can explain adipose tissue morphogenesis and that regeneration requires after injury the same mechanisms. We validated this hypothesis by means of a two-dimensional Individual Based Model (IBM) of interacting adipocytes and ECM fiber elements [135]. The model successfully generated regeneration or scar formation as functions of few key parameters, and seemed to indicate that the fate of injury outcome could be mainly due to ECM rigidity.

### **Modeling of morphogen diffusion in *Drosophila* oogenesis**

In collaboration with a team of developmental biologists of Rutgers University (Camden, New Jersey), we have built a model for the diffusion of the Gurken morphogen during *Drosophila* oogenesis, taking into account a wide variety of biological mechanisms such as diffusion of the morphogen, reactions of components of the EGFR signaling pathway, movement of the source of morphogen, shift of the overlying follicle cells and growth of the egg chamber. This model, together with a complete numerical code developed in Matlab, provides a tool to understand how each mechanism influences the signal distribution. The overall aim of the project is to use this tool to guide future experiments, and to understand what mechanisms contribute to the different distributions of signal among species.

### **Bacterial population growth**

We exploited all the methods developed to estimate the division rate of a population (see axis 3) to address a seminal question of biology: is it a size-sensing or a timing mechanism which triggers bacterial growth? In [138], we showed that a sizer model is robust and fits the data well. Several studies from other groups came at the same time, showing a renewed interest on a question dated back to Jacques Monod's PhD thesis (1941). Of special interest is the "adder" model, for which we are currently developing new estimation methods.

### **A quantitative high resolution computational mechanics cell model for growing and regenerating tissues**

---

<sup>0</sup>ravasio:hal-01245750, vedula:hal-01298859

Mathematical models are increasingly designed to guide experiments in biology, biotechnology, as well as to assist in medical decision-making. They are in particular important to understand emergent collective cell behavior. For this purpose, the models, despite still abstractions of reality, need to be quantitative in all aspects relevant for the question of interest. During the regeneration of liver after drug-induced depletion of hepatocytes surviving dividing and migrating hepatocytes must squeeze through a blood vessel network to fill the emerged lesions. Here, the cells' response to mechanical stress might significantly impact on the regeneration process. We developed a 3D high-resolution cell-based model integrating information from measurements in order to obtain a refined quantitative understanding of the cell-biomechanical impact on the closure of drug-induced lesions in liver. Our model represents each cell individually, constructed as a physically scalable network of viscoelastic elements, capable of mimicking realistic cell deformation and supplying information at subcellular scales. The cells have the capability to migrate, grow and divide, and infer the nature of their mechanical elements and their parameters from comparisons with optical stretcher experiments. Due to triangulation of the cell surface, interactions of cells with arbitrarily shaped (triangulated) structures such as blood vessels can be captured naturally. Comparing our simulations with those of so-called center-based models, in which cells have a rigid shape and forces are exerted between cell centers, we find that the migration forces a cell needs to exert on its environment to close a tissue lesion, is much smaller than predicted by center-based models. This effect is expected to be even more present in chronic liver disease, where tissue stiffens and excess collagen narrows pores for cells to squeeze through [44].

Main collaborators: Stefan Höhme, Univ. Leipzig; Josef Käs, Univ. Leipzig.

#### **Modeling the extracellular matrix in multicellular organization and liver regeneration:**

An important step has been undertaken to integrate an explicit model of collagen networks in liver and other tissues. The mechanical model of collagen fibers uses linear and rotational springs to represent collagen fibers and collagen network taking into account the stretching and bending energy of the collagen network. The model has been validated by direct comparison to experiments where a force has been exerted on a single collagen fiber, as well as to shear experiments. In a next step, this collagen model has been incorporated into the previous lobule model to simulate spatio-temporal pattern of the ECM in the lobule. One key objective is a model of fibrosis development, whereby fibrotic streets form. So far, no coherent model exist but a number of hypotheses that will be implemented and tested versus data.

Main collaborators: Steven Dooley, Seddik Hammad, Univ. Mannheim; JG Hengstler, IfADo.

#### **Models of flow in liver**

Also for the liver, a model for bile salts transport has been developed. The current hypothesis is that bile salt excreted by hepatocytes into bile canaliculi, are transported within canaliculi by convection through the canal of hering to the biliary ducts. In close iterations with experiments and by image based modeling running simulations directly in reconstructed 3D volume data sets, we test different alternative mechanisms of bile transport.

The blood flow model of the individual liver lobule, the smallest anatomical and functional repetitive unit of liver has been embedded in an electrical analogue model for the whole model hemodynamics to compute the impact of architectural changes at the lobule level as they occur after partial hepatectomy on the whole body hemodynamics. At the lobule level, the impact of capillary diameters on the flow have been studied under consideration of the hematocrit value (the volume fraction of red blood cells).

Collaborators: Chloé Aubert, UPMC, I. Vignon-Clementel, REO, Jan G. Henstler and Natiket Vartak, IfADo, Eric Vibert, Hopital Paul Brousse, Villejuif

#### **Liver development**

The deformable cell model is being used to establish a model of bile duct formation. Bile ducts form at the portal veins as a consequence of an interaction of cholangiocytes, aligning the mesenchyme of the portal vein, and hepatoblasts surrounding the portal veins. Hepatoblasts are a pre-stage of hepatocytes. Current biological hypotheses speculate that bile ducts may either emerge from proliferation of the hepatoblast layer that is in contact to cholangiocytes, leading to formation of a lumen by buckling, by attraction of water a positions



where cholangiocytes secrete mucin, or by apical contraction of hepatoblasts within the layer in contact to the cholangiocytes. The simulations are performed with the deformable cell model.

Collaborators: Frederic Lemaigre, De Duve Institut, Brussels, Anne Dubart-Kupperschmitt, Hopital Paul Brousse.

#### **Image processing, analysis and quantification of tissue microarchitecture**

At the interface between experiment and modeling we pursue a number of projects on image analysis and quantification. Such information in the past often served to generate hypotheses of the mechanisms underlying image sequences in time, which then could be turned in a mathematical model to verify, which hypotheses are sufficient to explain the image data. Several image analysis projects focus on the liver. (1.) Bile microinfarcts have been found to be initiated by rupture of the apical hepatocyte membrane leading to shunting from bile canaliculi to the blood capillaries. This is followed by massive increase of the immune cells as could be quantified by analysis of intravital micrographs. For every frame in the video, a binary mask that most likely resemble detected immune cells were obtained. The process started by applying suitable linear and non-linear filters to highlight structure of interest and remove noise. Morphological operations and blob analysis was finally utilized to locate and count the cells. With the help of an expert, the confusion matrix was finally established to assess the quality of the segmentation and the obtained results (Ghallab et. al., *J. Hepat.* 2018; <https://aasldpubs.onlinelibrary.wiley.com/doi/full/10.1002/hep.30213>. [14]) (2.) In a second project, the CYP – enzyme distribution is quantified after repetitive administration of CCl<sub>4</sub>. After single overdose of CCl<sub>4</sub>, the liver shows a peri-central liver lobule (smallest repetitive anatomical unit of liver) necrosis. This pattern changes in repetitive dosing and leads to chronic disease stages and sometimes eventually to either hepatocellular carcinoma [15] or acute-on-chronic liver failure (ACLF), a disease condition with often-lethal outcome. Data from whole-slide scans is analyzed to serve to develop a mathematical model of ACLF. (3.) A similar strategy is pursued for fibrosis formation through high fat diet both by image analysis of mouse and human data, aiming at a mathematical lobule model based on a deformable cell model. Here currently images are analyzed to quantify microarchitectural modifications as a consequence of Western Diet (a high fat diet generating a disease condition reminiscent of NAFLD in human. (4.) TiQuant-algorithms have been used to analyze micro- and macrovascular alterations in cirrhosis [125] and (5.) tissue modifications after PHx [82].

Main collaborators: Ahmed Ghallab, Jan G. Hengstler, IfADo; Ursula Klingmüller, DKFZ Heidelberg; Steven Dooley, Univ. Hospital Mannheim; Percy Knolle, Helmholtz Inst. Munich, Joachim Bode, Univ. Hospital Düsseldorf, Christian Trautwein, Univ. Hosp. Aachen; P. Seegers (Ghent University); Eric Vibert (Hopital Paul Brousse).

#### **Relating imaging on microscopic scales with imaging on macroscopic scales: From Diffusion-Weighted MRI Calibrated With Histological Data: an Example From Lung Cancer**

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumor treatment assessment, reflecting Brownian movement of water molecules in tissues. Since densely packed cells restrict molecule mobility, tumor tissues produce usually higher signal (less attenuated signal) on isotropic maps compared with normal tissues. However, no general quantitative relation between DWI data and the cell density has been established. In order to link low-resolution clinical cross-sectional data with high resolution histological information, we developed an image processing and analysis chain, which was used to study the correlation between the diffusion coefficient (D value) estimated from DWI and tumor cellularity from serial histological slides of a resected non-small cell lung cancer tumor. Color deconvolution followed by cell nuclei segmentation was performed on digitized histological images to determine local and cell-type specific 2d (two-dimensional) densities. From these, the 3d cell density was inferred by a model-based sampling technique, which is necessary for the calculation of local and global 3d tumor cell count. Next, DWI sequence information was overlaid with high resolution CT data and the resected histology using prominent anatomical hallmarks for co-registration of histology tissue blocks and non-invasive imaging modalities' data. The integration of cell numbers information and DWI data derived from different tumor areas revealed a clear negative correlation between cell density and D value. Importantly, spatial tumor cell density can be calculated based on DWI data. In summary, our results demonstrate that tumor cell count and heterogeneity can be predicted from DWI data, which may open new opportunities for personalized diagnosis



and therapy optimization [145]. The work of that paper has been further advanced to adapt the procedures for clinical use (in preparation).

### Collaborations

- Biological control of arboviroses: **Nicolas Vauchelet** (Université Paris 13); **Grégoire Nadin** (LJLL, Sorbonne Université); **Yannick Privat** (Université de Strasbourg); **D. Villela, C. Struchiner** (Fiocruz, Brazil); **Jorge Zubelli** (IMPA, Brazil); **Alain Rapaport** (INRA-Montpellier), **Y. Dumont** (CIRAD-Montpellier); **Ch. Schaerer, P. Pérez-Estigarribia** (UNA, Paraguay), **O. Vasilieva** (Universidad del Valle, Cali, Colombia), **D. Cardona-Salgado** (Universidad Autónoma de Occidente, Cali, Colombia).
- Protein aggregation in amyloid diseases: **Human Rezaei**'s team at Inra Jouy-en-Josas (France) and **W-F Xue**'s team in at university of Kent (Great Britain); **Tom Banks** at the North Carolina State University (USA) and **Philippe Moireau** (M3DISIM)
- Bacterial growth and division: **Lydia Robert**, Sorbonne Université (France)
- Liver research & toxicology: **JG. Hengstler** group (IfADo, Dortmund, Germany); **R. Gebhardt** (Univ. Leipzig); **U. Klingmueller** (DKFZ, Heidelberg); **Irène Vignon-Clementel** (Inria, REO)
- Wound healing: **Patrizia Bagnerini** (Genova, Numerical methods), **Benoît Ladoux** (Institut Jacques Monod et Mechanobiology Institute Singapore, Biophysics) and **Antonio Jacinto** (CEDOC, Lisbon, Biology and Medicine). (Adipose tissue regeneration) team of **L. Casteilla** (StromaLab, Toulouse)
- Diffusion of morphogen: Center for Computational and Integrative Biology, Rutgers University (Camden, New Jersey), joint work with Professor Nir Yakoby's Drosophila Laboratory
- Linking micro and macro-image information: Oliver Sedlaczek, Univ. and DKFZ Heidelberg, Kai Breuhahn, Univ. Heidelberg.

## MONC Project-Team

### 4. Application Domains

#### 4.1. Tumor growth monitoring and therapeutic evaluation

Each type of cancer is different and requires an adequate model. More specifically, we are currently working on the following diseases:

- Glioma (brain tumors) of various grades,
- Metastases to the lung, liver and brain from various organs,
- Soft-tissue sarcoma,
- Kidney cancer and its metastases,
- EGFR-mutated lung cancer.

In this context our application domains are:

- Image-driven patient-specific simulations of tumor growth and treatments,
- Parameter estimation and data assimilation of medical images.

#### 4.2. Biophysical therapies

- Modeling of electrochemotherapy on biological and clinical scales,
- Evaluation of radiotherapy and radiofrequency ablation.

#### 4.3. Experimental research in oncology

We use mathematical modeling as a tool to assist biological cancer research in testing mechanistic hypotheses and pharmacological research. More specifically, our applications include:

- Theoretical biology of the metastatic process: dynamics of a population of tumors in mutual interactions, dormancy, pre-metastatic and metastatic niche, quantification of metastatic potential and differential effects of anti-angiogenic therapies on primary tumor and metastases,
- Rational and personalized dosing and scheduling of anti-cancer drugs in monotherapy or in combination,
- Mathematical modeling of intra-tumor drug transport to optimize the design of nanoparticles

**NUMED Project-Team (section vide)**

## REO Project-Team

# 4. Application Domains

## 4.1. Blood flows

Cardiovascular diseases like atherosclerosis or aneurysms are a major cause of mortality. It is generally admitted that a better knowledge of local flow patterns could improve the treatment of these pathologies (although many other biophysical phenomena obviously take place in the development of such diseases). In particular, it has been known for years that the association of low wall shear stress and high oscillatory shear index give relevant indications to localize possible zones of atherosclerosis. It is also known that medical devices (graft or stent) perturb blood flows and may create local stresses favorable with atherogenesis. Numerical simulations of blood flows can give access to this local quantities and may therefore help to design new medical devices with less negative impacts. In the case of aneurysms, numerical simulations may help to predict possible zones of rupture and could therefore give a guide for treatment planning.

In clinical routine, many indices are used for diagnosis. For example, the size of a stenosis is estimated by a few measures of flow rate around the stenosis and by application of simple fluid mechanics rules. In some situations, for example in the case a sub-valvular stenosis, it is known that such indices often give false estimations. Numerical simulations may give indications to define new indices, simple enough to be used in clinical exams, but more precise than those currently used.

It is well-known that the arterial circulation and the heart (or more specifically the left ventricle) are strongly coupled. Modifications of arterial walls or blood flows may indeed affect the mechanical properties of the left ventricle. Numerical simulations of the arterial tree coupled to the heart model could shed light on this complex relationship.

One of the goals of the REO team is to provide various models and simulation tools of the cardiovascular system. The scaling of these models will be adapted to the application in mind: low resolution for modeling the global circulation, high resolution for modeling a small portion of vessel.

## 4.2. Respiratory tracts

Breathing, or “external” respiration (“internal” respiration corresponds to cellular respiration) involves gas transport through the respiratory tract with its visible ends, nose and mouth. Air streams then from the pharynx down to the trachea. Food and drink entry into the trachea is usually prevented by the larynx structure (epiglottis). The trachea extends from the neck into the thorax, where it divides into right and left main bronchi, which enter the corresponding lungs (the left being smaller to accommodate the heart). Inhaled air is then convected in the bronchus tree which ends in alveoli, where gaseous exchange occurs. Surfactant reduces the surface tension on the alveolus wall, allowing them to expand. Gaseous exchange relies on simple diffusion on a large surface area over a short path between the alveolus and the blood capillary under concentration gradients between alveolar air and blood. The lungs are divided into lobes (three on the right, two on the left) supplied by lobar bronchi. Each lobe of the lung is further divided into segments (ten segments of the right lung and eight of the left). Inhaled air contains dust and debris, which must be filtered, if possible, before they reach the alveoli. The tracheobronchial tree is lined by a layer of sticky mucus, secreted by the epithelium. Particles which hit the side wall of the tract are trapped in this mucus. Cilia on the epithelial cells move the mucous continually towards the nose and mouth.

Each lung is enclosed in a space bounded below by the diaphragm and laterally by the chest wall and the mediastinum. The air movement is achieved by alternately increasing and decreasing the chest pressure (and volume). When the airspace transmural pressure rises, air is sucked in. When it decreases, airspaces collapse and air is expelled. Each lung is surrounded by a pleural cavity, except at its hilum where the inner pleura give birth to the outer pleura. The pleural layers slide over each other. The tidal volume is nearly equal to 500 ml.

The lungs may fail to maintain an adequate supply of air. In premature infants surfactant is not yet active. Accidental inhalation of liquid or solid and airway infection may occur. Chronic obstructive lung diseases and lung cancers are frequent pathologies and among the three first death causes in France.

One of the goals of REO team in the ventilation field is to visualize the airways (virtual endoscopy) and simulate flow in image-based 3D models of the upper airways (nose, pharynx, larynx) and the first generations of the tracheobronchial tree (trachea is generation 0), whereas simple models of the small bronchi and alveoli are used (reduced-basis element method, fractal homogenization, multiphysics homogenization, lumped parameter models), in order to provide the flow distribution within the lung segments.

### **4.3. Cardiac electrophysiology**

The purpose is to simulate the propagation of the action potential in the heart. A lot of works has already been devoted to this topic in the literature (see *e.g.* [40], [44], [43] and the references therein), nevertheless there are only very few studies showing realistic electrocardiograms obtained from partial differential equations models. Our goal is to find a compromise between two opposite requirements: on the one hand, we want to use predictive models, and therefore models based on physiology, on the other hand, we want to use models simple enough to be parametrized (in view of patient-specific simulations). One of the goal is to use our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macs/M3disym project-team, we are interested in the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology, in particular the simulation of experimental measurement of the field potential of cardiac stem cells in multi-electrode arrays.

## SISTM Project-Team

### 4. Application Domains

#### 4.1. Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from "bench side to bedside") and vice versa, a process often called "Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [45], "Systems vaccinology" [42], "Systems medicine" [34]. From the statistician point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

#### 4.2. The case of HIV immunology

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [38]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [33]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

#### 4.3. The case of Ebola vaccine development

In response to the recent outbreak of Ebola virus disease in West Africa, the clinical development of some candidate to Ebola vaccine has been accelerated. Several vectors, mostly encoding glycoprotein of the virus, were tested in Phase I-II studies in order to assess their safety and immunogenicity. One of the main question of interest there is the antibody response induced by vaccination, as some non-human primates studies have shown protection against the virus when antibody levels were high enough. Although bridging studies still have to be developed, antibodies are thus considered as a criterium of interest. The challenge is then to evaluate the durability of the antibody response, whether it be at an individual or population level, in order to evaluate the impact of a vaccine strategy in case of an epidemic. Moreover, we are interested in the factors associated to this antibody response, and even more the other immune markers (from both innate and adaptative immune response) able to predict antibody levels. As those relationship are non-linear, sophisticated statistical and mathematical methods are developed in order to address these questions. A systems medicine approach using multidimensional immunogenicity data from clinical trials and statistical models can help to understand vaccine mechanisms and improve the selection of optimised vaccine strategies for clinical trials.

## XPOP Project-Team

# 4. Application Domains

## 4.1. Precision medicine and pharmacogenomics

Pharmacogenomics involves using an individual's genome to determine whether or not a particular therapy, or dose of therapy, will be effective. Indeed, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.

Currently, in the traditional population approach, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient.

The challenge is to determine which genetic covariates are associated to some PKPD parameters and/or implicated in patient responses to a given drug.

Another problem encountered is the dependence of genes, as indeed, gene expression is a highly regulated process. In cases where the explanatory variables (genomic variants) are correlated, Lasso-type methods for model selection are thwarted.

There is therefore a clear need for new methods and algorithms for the estimation, validation and selection of mixed effects models adapted to the problems of genomic medicine.

A target application of this project concerns the lung cancer.

EGFR (Epidermal Growth Factor Receptor) is a cell surface protein that binds to epidermal growth factor. We know that deregulation of the downstream signaling pathway of EGFR is involved in the development of lung cancers and several gene mutations responsible for this deregulation are known.

Our objective is to identify the variants responsible for the disruption of this pathway using a modelling approach. The data that should be available for developing such model are ERK (Extracellular signal-regulated kinases) phosphorylation time series, obtained from different genetic profiles.

The model that we aim to develop will describe the relationship between the parameters of the pathway and the genomic covariates, i.e. the genetic profile. Variants related to the pathway include: variants that modify the affinity binding of ligands to receptors, variants that modify the total amount of protein, variants that affect the catalytic site,...

## 4.2. Oncology

In cancer, the most dreadful event is the formation of metastases that disseminate tumor cells throughout the organism. Cutaneous melanoma is a cancer, where the primary tumor can easily be removed by surgery. However, this cancer is of poor prognosis; because melanomas metastasize often and rapidly. Many melanomas arise from excessive exposure to mutagenic UV from the sun or sunbeds. As a consequence, the mutational burden of melanomas is generally high

RAC1 encodes a small GTPase that induces cell cycle progression and migration of melanoblasts during embryonic development. Patients with the recurrent P29S mutation of RAC1 have 3-fold increased odds at having regional lymph nodes invaded at the time of diagnosis. RAC1 is unlikely to be a good therapeutic target, since a potential inhibitor that would block its catalytic activity, would also lock it into the active GTP-bound state. This project thus investigates the possibility of targeting the signaling pathway downstream of RAC1.

XPOP is mainly involved in Task 1 of the project: *Identifying deregulations and mutations of the ARP2/3 pathway in melanoma patients.*

Association of over-expression or down-regulation of each marker with poor prognosis in terms of invasion of regional lymph nodes, metastases and survival, will be examined using classical univariate and multivariate analysis. We will then develop specific statistical models for survival analysis in order to associate prognosis factors to each composition of complexes. Indeed, one has to implement the further constraint that each subunit has to be contributed by one of several paralogous subunits. An original method previously developed by XPOP has already been successfully applied to WAVE complex data in breast cancer.

The developed models will be rendered user-friendly through a dedicated Rsoftware package.

This project can represent a significant step forward in precision medicine of the cutaneous melanoma.

### 4.3. Hemodialysis

Hemodialysis is a process for removing waste and excess water from the blood and is used primarily as an artificial replacement for lost kidney function in people with kidney failure. Side effects caused by removing too much fluid and/or removing fluid too rapidly include low blood pressure, fatigue, chest pains, leg-cramps, nausea and headaches.

Nephrologists must therefore correctly assess the hydration status in chronic hemodialysis patients and consider fluid overload effects when prescribing dialysis, according to a new study.

The fluid overload biomarker, B-type natriuretic peptide (BNP) is an important component of managing patients with kidney disease. Indeed, it is believed that each dialysis patient will have an ideal or "dry" BNP level which will accurately and reproducibly reflect their optimal fluid status.

The objective of this study is to develop a model for the BNP and the hydration status using individual information (age, sex, ethnicity, systolic blood pressure, BMI, coronary heart disease history, ...).

The impact will be significant if the method succeeds. Indeed, it will be possible for the nephrologists to use this model for monitoring individually each treatment, in order to avoid risks of hypotension (low BNP) or overweight (high BNP).

### 4.4. Intracellular processes

Significant cell-to-cell heterogeneity is ubiquitously-observed in isogenic cell populations. Cells respond differently to a same stimulation. For example, accounting for such heterogeneity is essential to quantitatively understand why some bacteria survive antibiotic treatments, some cancer cells escape drug-induced suicide, stem cell do not differentiate, or some cells are not infected by pathogens.

The origins of the variability of biological processes and phenotypes are multifarious. Indeed, the observed heterogeneity of cell responses to a common stimulus can originate from differences in cell phenotypes (age, cell size, ribosome and transcription factor concentrations, etc), from spatio-temporal variations of the cell environments and from the intrinsic randomness of biochemical reactions. From systems and synthetic biology perspectives, understanding the exact contributions of these different sources of heterogeneity on the variability of cell responses is a central question.

The main ambition of this project is to propose a paradigm change in the quantitative modelling of cellular processes by shifting from mean-cell models to single-cell and population models. The main contribution of XPOP focuses on methodological developments for mixed-effects model identification in the context of growing cell populations.

- Mixed-effects models usually consider an homogeneous population of independent individuals. This assumption does not hold when the population of cells (i.e. the statistical individuals) consists of several generations of dividing cells. We then need to account for inheritance of single-cell parameters in this population. More precisely, the problem is to attribute the new state and parameter values to newborn cells given (the current estimated values for) the mother.



- The mixed-effects modelling framework corresponds to a strong assumption: differences between cells are static in time (ie, cell-specific parameters have fixed values). However, it is likely that for any given cell, ribosome levels slowly vary across time, since like any other protein, ribosomes are produced in a stochastic manner. We will therefore extend our modelling framework so as to account for the possible random fluctuations of parameter values in individual cells. Extensions based on stochastic differential equations will be investigated.
- Identifiability is a fundamental prerequisite for model identification and is also closely connected to optimal experimental design. We will derive criteria for theoretical identifiability, in which different parameter values lead to non-identical probability distributions, and for structural identifiability, which concerns the algebraic properties of the structural model, i.e. the ODE system. We will then address the problem of practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. An interesting problem is whether accounting for lineage effects can help practical identifiability of the parameters of the individuals in presence of measurement and biological noise.

## 4.5. Population pharmacometrics

Pharmacometrics involves the analysis and interpretation of data produced in pre-clinical and clinical trials. Population pharmacokinetics studies the variability in drug exposure for clinically safe and effective doses by focusing on identification of patient characteristics which significantly affect or are highly correlated with this variability. Disease progress modeling uses mathematical models to describe, explain, investigate and predict the changes in disease status as a function of time. A disease progress model incorporates functions describing natural disease progression and drug action.

The model based drug development (MBDD) approach establishes quantitative targets for each development step and optimizes the design of each study to meet the target. Optimizing study design requires simulations, which in turn require models. In order to arrive at a meaningful design, mechanisms need to be understood and correctly represented in the mathematical model. Furthermore, the model has to be predictive for future studies. This requirement precludes all purely empirical modeling; instead, models have to be mechanistic.

In particular, physiologically based pharmacokinetic models attempt to mathematically transcribe anatomical, physiological, physical, and chemical descriptions of phenomena involved in the ADME (Absorption - Distribution - Metabolism - Elimination) processes. A system of ordinary differential equations for the quantity of substance in each compartment involves parameters representing blood flow, pulmonary ventilation rate, organ volume, etc.

The ability to describe variability in pharmacometrics model is essential. The nonlinear mixed-effects modeling approach does this by combining the structural model component (the ODE system) with a statistical model, describing the distribution of the parameters between subjects and within subjects, as well as quantifying the unexplained or residual variability within subjects.

The objective of XPOP is to develop new methods for models defined by a very large ODE system, a large number of parameters and a large number of covariates. Contributions of XPOP in this domain are mainly methodological and there is no privileged therapeutic application at this stage.

However, it is expected that these new methods will be implemented in software tools, including MONOLIX and Rpackages for practical use.

## 4.6. Mass spectrometry

One of the main recent developments in analytical chemistry is the rapid democratization of high-resolution mass spectrometers. These instruments produce extremely complex mass spectra, which can include several hundred thousand ions when analyzing complex samples. The analysis of complex matrices (biological, agri-food, cosmetic, pharmaceutical, environmental, etc.) is precisely one of the major analytical challenges of this new century. Academic and industrial researchers are particularly interested in trying to quickly and effectively establish the chemical consequences of an event on a complex matrix. This may include, for

example, searching for pesticide degradation products and metabolites in fruits and vegetables, photoproducts of active ingredients in a cosmetic emulsion exposed to UV rays or chlorination products of biocides in hospital effluents. The main difficulty of this type of analysis is based on the high spatial and temporal variability of the samples, which is in addition to the experimental uncertainties inherent in any measurement and requires a large number of samples and analyses to be carried out and computerized data processing (up to 16 million per mass spectrum).

A collaboration between XPOP and the Molecular Chemistry Laboratory (LCM) of the Ecole Polytechnique began in 2018. Our objective is to develop new methods for the statistical analysis of mass spectrometry data.