

## EVALUATION REPORT OF THE UNIT

Cancer et génome : bioinformatique, biostatistiques et épidémiologie d'un système complexe

UNDER THE SUPERVISION OF THE  
FOLLOWING ESTABLISHMENTS AND  
ORGANISMS:

Institut Curie

Institut national de la santé et de la recherche  
médicale - Inserm

Université Paris sciences et lettres – Université PSL

Mines Paris – université Paris Sciences & Lettres -  
Mines Paris-PSL

---

**EVALUATION CAMPAIGN 2023-2024**  
GROUP D

Report published on June, 19 2024



In the name of the expert committee :

Jacques Colinge, Chairman of the committee

For the Hcéres :

Stéphane Le Bouler, acting president

Pursuant to Articles R. 114-15 and R. 114-10 of the Research Code, the evaluation reports drawn up by the expert committees are signed by the chairs of these committees and countersigned by the president of Hcéres.

To make the document easier to read, the names used in this report to designate functions, professions or responsibilities (expert, researcher, teacher-researcher, professor, lecturer, engineer, technician, director, doctoral student, etc.) are used in a generic sense and have a neutral value.

This report is the result of the unit's evaluation by the expert committee, the composition of which is specified below. The appreciations it contains are the expression of the independent and collegial deliberation of this committee. The numbers in this report are the certified exact data extracted from the deposited files by the supervising body on behalf of the unit.

## MEMBERS OF THE EXPERT COMMITTEE

### **Chairperson:**

Mr Jacques Colinge, Université de Montpellier

### **Experts:**

Ms Jessica Andreani, Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Paris

Ms Emmanuelle Bouzigon, Inserm, Paris

Ms Marie-Agnès Dillies, Institut Pasteur Paris

Mr Emmanuel Grenier, Université de Lyon

Mr Jérôme Lambert, Université Paris Cité

Mr Yves Moreau, KU Leuven Belgique

## HCÉRES REPRESENTATIVE

Mr Arnaud Seigneurin

## REPRESENTATIVES OF SUPERVISING INSTITUTIONS AND BODIES

Mr Julien Haccoun, Mines Paris-PSL

Ms Tatiana Malherbe, Institut Curie

Ms Anaïs Solle, Inserm

Mr Arnaud Tourin, Université PSL

## CHARACTERISATION OF THE UNIT

- Name: Computational Oncology
- Acronym: U900-CBIO
- Label and number: U900
- Composition of the executive team: Emmanuel Barillot

## SCIENTIFIC PANELS OF THE UNIT

SVE Sciences du vivant et environnement  
SVE7 Prévention, diagnostic et traitement des maladies humaines

## THEMES OF THE UNIT

The U900-CBIO Computational Oncology unit is devoted to the analysis and modeling of cancer-related data. It is comprised of six research teams, and a large bioinformatic core that constitutes the seventh team. Each research team has a strong methodological development activity to support its applications.

Team 1 investigates genetic epidemiology in familial cancers with a strong focus on breast and ovary tumors. Team 2 develops mathematical models of cancer regulation and spatial organization to understand tumor complexity, and find potential new therapeutic options in various carcinomas and sarcomas. Team 3 is interested in adult gliomas and pediatric brain tumors through data integration, with an emphasis on spatial heterogeneity and IDH-mutant gliomas. Team 4 develops machine learning methods to model different aspects of biological systems, in particular subcellular RNA localization, gene exchange in DNA, biomarkers. Team 4 is also interested in drug discovery through virtual screening. Team 5 combines multi-omics and multiscale data to unravel novel combined therapies and chronotherapy regimens. New tools of quantitative systems pharmacology based on pharmacodynamics and pharmacokinetics models are developed for this purpose. Team 6 develops statistical methods for precision medicine that rely on the proper use of biomarkers. They also investigate the evaluation of statistical methods and medical practice in medical trials and routine.

The large core facility (20 staff) named CUBIC covers a broad range of expertise supporting IC research, from software engineering to biostatistics, through bioinformatics, data management, and high-performance computing.

## HISTORIC AND GEOGRAPHICAL LOCATION OF THE UNIT

The emergence of high-throughput genomic technologies, and the need for computational expertise to deal with the data generated, led the IC to establish in 2002 a bioinformatic core facility. This initiated the genesis of the current structure U900-CBIO. Considering the development of computational biology, the increased level of required skills as well as their growing diversity, a stronger commitment was envisioned. In 2008, a partnership between IC, Inserm, and Mines Paris - PSL enabled the creation of a dedicated unit named U900-CBIO. Since its inception in 2008, the unit has been evaluated twice by Hcéres (2014 & 2019). The sole core facility of 2002 evolved into a structure with three research teams and a core at the beginning of the previous mandate (2019), to reach end 2022 an organization with six research teams and a large core of 20 staff; total staff 97.

Currently, U900-CBIO teams are located at four sites: teams 3, 5, and 6 are located at IC Saint-Cloud site; team 2, 7 (core) and half of 4 are at IC main site, rue Pierre et Marie Curie; the second half of team 4 is at Mines Paris - PSL site; team 1 is at rented building by IC, bld Saint-Michel. The last three sites are at walking distance, while the Saint-Cloud site requires 45 min public transport to be reached from the others. In 2025, team 1 will move to the Saint-Cloud site, whereas in 2030-2031 teams 2 and 4 will move to the future Parisanté Campus Val-de-Grâce located close to their current premises.

## RESEARCH ENVIRONMENT OF THE UNIT

The computational biology unit is formally a IC-Inserm research unit (U900) and a Mines Paris - PSL research center (center for computational biology, CBIO). At the moment of this evaluation, U900 teams are 1, 2, 6, and 7, whereas team 4 is CBIO. Team 6 was created by IC in 2016, but not endorsed by Inserm at previous evaluation. During the current mandate, two junior teams were created: 3 (F Cavalli, 2020) and 5 (A Ballesta, 2021). They both obtained ATIP starting grants. U900-CBIO is part of the IC comprehensive research center. As described above, all the teams are hosted in IC premises with the exception of team 4, one half of which is located at Mines Paris premises.

U900-CBIO supports IC and Inserm in the construction of many partnerships, e.g., in EQUIPEX project for NGS sequencing or high-performance computing (HPC), IC single cell Initiative, IC involvement in France Médecine Génomique, the construction of the Comprehensive Cancer Center (SIRIC), RHU CASSIOPEIA & epCART, IHU Women's Cancers. All these large projects rely heavily on expertise in bioinformatics, computational biology,

biostatistics and AI brought by U900-CBIO.

Furthermore, teams 2 & 4 are involved in Parisanté Campus, a joint effort by Inserm, PSL University, Inria, Health Data Hub, and Agence Numérique en Santé to gather forces in numerical health, public and private. U900-CBIO is also involved in PRAIRIE, Paris Artificial Intelligence Research institute, another private-public effort to foster education, research and applications in AI. Teams 2 & 4 obtained a total of four chairs at PRAIRIE.

## UNIT WORKFORCE: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	3
Maîtres de conférences et assimilés	3
Directeurs de recherche et assimilés	2
Chargés de recherche et assimilés	4
Personnels d'appui à la recherche	36
<b>Sous-total personnels permanents en activité</b>	<b>48</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	4
Personnels d'appui non permanents	0
Post-doctorants	7
Doctorants	30
<b>Sous-total personnels non permanents en activité</b>	<b>41</b>
<b>Total personnels</b>	<b>89</b>

## DISTRIBUTION OF THE UNIT'S PERMANENTS BY EMPLOYER: in physical persons at 31/12/2022. Non-tutorship employers are grouped under the heading "others".

Nom de l'employeur	EC	C	PAR
INST CURIE	4	1	29
Inserm	0	4	4
AUTRES	1	0	3
MINES PARIS-PSL	1	1	0
<b>Total personnels</b>	<b>6</b>	<b>6</b>	<b>36</b>

## GLOBAL ASSESSMENT

Institut Curie Computational Oncology unit U900-CBIO is a joint initiative of IC-Inserm (all teams but 4), and Mines Paris - PSL (team 4). It is located at three IC sites and one Mines Paris - PSL site. It occupies a rare position in the bioinformatics/biostatistics French landscape. It is a multi-team computational institute that is embedded with fundamental biology research and clinical operations. It also includes a bioinformatics platform, which enables to engage in collaborations with biologists and clinicians at all the necessary levels, from infrastructure or routine support to pure methodological developments.

IC and Paris as a whole are strong poles of cancer research, which obviously creates numerous opportunities in collaborative work and fund raising. U900-CBIO has been remarkably efficient to capture these opportunities to both develop its own activities, but also give back and help their main hosting institution (IC), and to a certain extent the Paris scene, developing their research programs. External funding (18M€) represented >90% of the unit resources, with many national (10x Inca, 5x ITMO, 8x ANR), European (6x H2020, 1x ERA-NET), and industrial grants. The unit is an actor of several structuring efforts (France Médecine Génomique, SIRIC, RHU CASSIOPEIA

& epCART, IHU Women's Cancers, ParisSanté Campus, PRAIRIE). This outstanding success is especially relevant since U900-CBIO was precisely created to help IC to integrate molecular biology high-throughput genomic technologies.

In terms of science, productivity was outstanding with 540 research papers published, more than 30% with corresponding authorship (Nat Genetics, Genome Biol, Nat Comm, Cell Rep, etc.). The different research themes of all the teams complement each other nicely to cover many essential computational needs of cancer research, from genetic association studies to innovations in the statistical analysis of clinical trial data. Spatial and single-cell biology that are major game changers in the current cancer research have been identified as main targets and are the objects of multi-team collaborations within U900-CBIO. The same is true for AI applications.

The unit interacts actively with the non-academic society. A number of CIFRE PhD fellowships were obtained with different pharma and biotech partners. A total of eighteen research contracts involving private companies were acquired (Bristol-Myer Squibb, 3 x Servier, 2 x SANOFI, 2 x Intel). Moreover, and most importantly, direct financial support was found at private actors such as Intel to provide a powerful compute cluster serving IC as a whole. Many initiatives were also undertaken towards the general public through all sorts of scientific vulgarization channels (radio broadcasts, books for children, journées de la science, etc.). The media and industry presence of the unit is excellent.

To finish, through the hiring of two new junior teams, but also with the new leader of team 1 (F. Lesueur) who has dual experimental-computational expertise, U900-CBIO will explore hosting teams that are not purely computational, or teams who have the ability to generate part if not all the data they need. This is a new turn, which sounds attractive and timely. Biologists with solid data analysis skills or bioinformaticians with a hard science background that run a lab in addition to their computational programs has become a trend. It brings new potential to the unit provided the latter does not reduce its efficiency on its primary missions that is to foster IC research.

The committee wants to unanimously underline the remarkable role of U900 director, Dr. E Barillot, in the construction of this outstanding unit. He managed to transform an initial support-driven initiative in 2002 into a rich, cutting-edge research-driven organization, while still providing basic support, and letting U900-CBIO researchers develop and shine for themselves as well.

## DETAILED EVALUATION OF THE UNIT

### A - CONSIDERATION OF THE RECOMMENDATIONS IN THE PREVIOUS REPORT

U900-CBIO has implemented convincing actions in response to recommendations formulated in previous Hcéres report. They essentially addressed all the points under their control. Some issues dependent on IC or Mines Paris - PSL were not addressed (space, engineer recruitments).

In a first set of recommendations, U900-CBIO was (A) suggested to increase the number of PhD students and make them publish in top journals. Teams 1, 2, and 4 increased their numbers of PhD students, and most published several papers. In team 2, best papers were often issued from the work of PhD students. (B) U900-CBIO was recommended to maintain links with industry. Six CIFRE PhD theses were started with biotech or pharma companies (SANOFI for teams 4 & 6); two PhD were directly funded by a pharma (Janssen to team 4) or biotech. Participation in clinical trials (team 6 with MSD) and additional operations such as Intel that provided a compute cluster for >300k€. A further recommendation (C) was to clarify the policy for deciding which pieces of software are maintained and potentially shared. U900-CBIO has established standard practice to standardize software development and make it available using containers. This facilitates maintenance by design. What is actually maintained is decided based on utility and audience.

A second set addressed the unit's organization: (A) Need to clarify the diverse missions of the core team 7. This team is focused on new technology support, pipeline development and maintenance, training, and data analysis tasks that do not need extensive new development. (B) Development of transversal axes between teams. This has been installed, e.g., for spatial transcriptomics through PhD co-supervision and shared seminars. (C) Promotion of women. The unit is supportive to women, many of which are very successful and rewarded accordingly. (D) Adequate planning of annual interviews. These interviews are mandatory at U900-CBIO and subject to written report to HR. (E) Lack of space and unit split at multiple locations. Constraints are such that a single location is not foreseeable. Additional space is planned nonetheless.

A third set was related to scientific strategy and projects. (A) Increase synergies between teams. Junior Team 6 has established many internal collaborations and thus gets support and guidance. Beyond this example, several other active collaborations are in place through expertise sharing or PhD co-supervision mostly. (B) Increase the number of researchers. This has been done first with two new, junior teams, and, second, through the recruitment of three researchers and three professors. Another three positions are still open. (C) Engineers to turn prototype software into sustainable packages. Supervisory institutions (IC & Mines Paris - PSL) did not address this obvious need, which team 2 solved on grant money. The aforementioned standards for code production and release also help. Active maintenance and development depend on internal needs. (D) Increased role in developing bioinformatics pipelines for IC as a whole. The core team was involved in the development of >20 such pipelines used on IC technology platforms mostly.

### B - EVALUATION AREAS

#### EVALUATION AREA 1: PROFILE, RESOURCES AND ORGANISATION OF THE UNIT

##### Assessment on the scientific objectives of the unit

The unit has a dual function of direct support but also co-construction of ambitious, computationally intensive cancer projects with IC and U900-CBIO partners at other institutions. From this perspective, the scientific objectives are excellent, covering many essential or extremely useful elements of cancer data analysis. There are a number of cross team projects (e.g., integration of multi-modal data, network-driven GWAS, clinical impact of mechanistic models), but no specific large common scientific goals.

##### Assessment on the unit's resources

U900-CBIO has managed to raise significant national and European funding. It is involved in important initiatives to support digital health (PariSanté Campus) and AI (PRAIRIE), which also provide financial support. Solid compute infrastructure was provided by Intel. Resources are outstanding. Administrative support is

limited though.

### Assessment on the functioning of the unit

Despite being located at multiple sites, U900-CBIO harbors several shared projects between teams. They have unit seminar every two weeks, a unit retreat every three years, a protocol to introduce each new comer personally to every team on her first day of work, and several cross team and hence sites thematic clubs. Given the space difficult conditions, the functioning as a whole is excellent.

#### *1/ The unit has set itself relevant scientific objectives.*

##### Strengths and possibilities linked to the context

The different senior teams and the core unit have established themselves as strong players in their respective domains, which are all highly relevant. Their network of collaborations is impressive and ensures continued access to data and funding. The junior teams have secured solid initial funding and the rich environment of U900-CBIO should help them to achieve their full potential. Shared standards for software development.

##### Weaknesses and risks linked to the context

Some teams develop project in response to experimental or clinical group needs predominantly, which is in line with the mission of the unit, but could be regarded as a limitation. The set of skills offered by the different teams cover a broad range of what is needed in modern cancer research, but most of the efforts aim at the excellence of each team independently. There are several common projects, though with rather technical or specific objectives. No big common objective.

#### *2/ The unit has resources that are suited to its activity profile and research environment and mobilises them.*

##### Strengths and possibilities linked to the context

Curie and Inserm financial endowment only represent 9.4% of U900 funding, 90.7% was raised from third-party at national and European levels. Roughly, 19M€ were raised in total during the previous mandate, 13.5M€ from national funding bodies (INCa, Fondation ARC, ANR, Ligue contre le cancer), >4M€ from the European Union, including UK, and almost 1.5M€ were obtained from industrial partners. Fund raising is thus outstanding.

Access to data, both clinical and molecular, is outstanding as well as being part of IC, but also thanks to the many collaborators of U900-CBIO teams at other institutions in France and abroad.

The unit has access to two compute clusters (Curie, 3000 cores, 5.5PB storage, and Mines with less CPUs but many GPUs). There are plans to access additional servers in the future (Jean Zay super computer thanks to PRAIRIE involvement, and MUDIS4LS national Equipex+ project).

##### Weaknesses and risks linked to the context

More intellectual property could be generated (2 filed patents only) to potentially generate revenues in addition to PhD fellowships.

Limited support to administration and technical positions by IC/Inserm/Mines (2 admin personal currently, one retiring soon with no clear replacement plan). This is a potential burden to PIs that must assume excessive administrative tasks.

Three teams (2, 5, 7) raise all the European money, the other teams have no European projects. Same differences with respect to industrial money raised by teams 2, 4, and 7 only if one disregards CIFRE PhD fellowship that consist in public money primarily.

#### *3/ The unit's practices comply with the rules and directives laid down by its supervisory bodies in terms of human resources management, safety,*



*environment, ethical protocols and protection of data and scientific heritage.*

### Strengths and possibilities linked to the context

Being integrated in IC, U900-CBIO follows IC rules and relies on its administration for many aspects, which ensures professional management and compliance with a number of rules. An example is IC gender equality, diversity and inclusion plan (GEDIP).

Protection of data is of special relevance at U900-CBIO. CUBIC (team 7) has developed an in-house secure IT system to control access to compute resources and data. Solutions are being explored to reduce the climate impact of computing through better energy management.

### Weaknesses and risks linked to the context

Psychological risk due to strong competition in cancer research, which can impact individuals.

## EVALUATION AREA 2: ATTRACTIVENESS

### Assessment on the attractiveness of the unit

Overall, U900-CBIO attractivity is outstanding for young researchers, engineers or students who are interested in pursuing or starting a computational life science career. The unit conducts research covering both purely methodological aspects and applications such that each researcher can potentially find his / her ideal spot. Available resources are also outstanding.

- 1/ The unit has an attractive scientific reputation and is part of the European research area.*
- 2/ The unit is attractive because for the quality of its staff support policy.*
- 3/ The unit is attractive through its success in competitive calls for projects.*
- 4/ The unit is attractive for the quality of its major equipment and technical skills.*

### Strengths and possibilities linked to the context for the four references above

1/ The unit has an attractive scientific reputation and contributes to the construction of the European research area.

U900-CBIO is one of the very few French, and one of the few European computational units, which is strongly linked to fundamental and translational cancer research with direct links to the clinic. It is thus a flagship unit for bioinformatics/biostatistics activities at the interface between pure methodological development and full commitment to applications. Over the years, its researchers managed to exploit this opportunity to achieve outstanding attractivity.

Since 2017, they have given >300 talks at national and international conferences and workshops. Several prizes were received such as Chloé Azencott (young AI woman engineer award, 2021).

They have organized a large number of conferences some of which are recurrent events, e.g., the yearly SysMod COSI 2016-2021 (L Calzone, team 2), BioNetVisa workshop 2017-2021 (I Kuperstein, E Barillot, A Zinovyev, T Walter, teams 2 & 4), C Azencott organized several machine learning meetings (Prague 2017, Paris 2020 & 2021), X Paoletti & A Latouche organized EPICLIN 2022 the meeting of the French cancer centers statisticians.

U900-CBIO members sit in a number of steering committees, for example, Institut Français de Bioinformatique (E Barillot), GDR ImaBio executive board (T Walter), statisticians from the French cancer centers (X Paoletti), clinical trials committee of the European consortium Innovative Therapies for Children with Cancer (ITCC) up to 2019 (X Paoletti).

Lastly, membership at important institutions also involves several teams or individual researchers. For instance, CUBIC (team 7) is a member of Institut Français de Bioinformatique, and of CoreForLife, an Excellence Alliance of Life Science Core Facilities in Europe for ten years. The Walter team is part of France Bioimaging (FBI), a

national infrastructure for microscopy, aiming at providing access to cutting edge microscopy to the scientific community in France and Europe. T.Walter, E.Barillot, C.Azencott and A.Zinovyev hold chairs at PRAIRIE.

2/ The unit is attractive for the quality of its staff hosting policy

IC has been awarded the "HR Excellence in Research label" by the European Commission for its European human resources strategy for researchers. This label acknowledges continuous and active improvement processes to offer a favorable and stimulating working environment. Beyond IC standards, U900-CBIO has set up a unique new employee integration process. New members are presented informally to all teams at the day of their arrival, they receive a welcome booklet to provide them with all the necessary information about the functioning of the institute, and of the unit. Several activities are offered to new PhD students and postdocs to help them integrating the unit.

Despite limited space, U900-CBIO managed to host visitors on a regular basis including for long periods. Examples of such hosts were Maximiliano Ribeiro-Guerra, Senior Visiting Professor (Federal University of Juiz de Fora, MG, Brazil) (10/2016-09/2017); David Goldgar, Senior visiting Professor (Huntsman Cancer Institute, University of Utah, USA) , recipient of the Mayent-Rothschild fellowship from Institut Curie for sabbaticals (09/2019 – 12/2019); Justina Zvirblykte , Visiting PhD student from Vilnius University (EC funding from her lab), 2022; Hanaé Carrié, Visiting PhD from the National University of Singapore, 2021.

Scientific integrity is taken very seriously at IC, and researchers are introduced and trained in this notion thoroughly, which participate in their continuous training.

3/ The unit is attractive because of the recognition gained through its success in competitive calls for projects  
With extra mural funding reaching 90.7% of its total budget (not including permanent staff salaries), U900-CBIO managed to be extremely successful in research grant applications. Europe: 6 x H2020 and 1 ERA-NET. France: 10 x INCa (including 3 PLBIO, 1 PRT-K, 1 SHS), 5 x ITMO, 8 x ANR, 1 x Inserm. Region: 4 x PRAIRIE, 2 x Cancéropôle. Charities: 18 grants (4 x LNCC Paris, 3 x Fondation ARC, 2 x Fondation Maladies rares, 2 x FRM). Private companies: 18 contracts (including Bristol-Myer Squibb, 3 x Servier, 2 x SANOFI, 2 x Intel).

4/ The unit is attractive for the quality of its major equipment and technological skills

The compute infrastructure and data management set up at IC are excellent. Access to additional compute clusters for additional power or specific computations (GPUs) is possible at Mines cluster and soon at two additional clusters (thanks to PariSanté Campus partnership and future equipex+ MUDIS4LS). Moreover, the establishment of standard guidelines as well as well-maintained bioinformatics pipelines provide bases for starting new projects or increase productivity.

Team 1 has established a technology platform to conduct investigations in genetics and epidemiology (PIGE).

Weaknesses and risks linked to the context for the four references above

U900-CBIO is active in an area where the recruitment of qualified students and postdocs in computer science, machine learning, applied statistics, and bioinformatics is necessary to implement projects successfully. Strong competition with the economic sector can be an issue in attracting the best – or good enough – candidates. Although all the teams raised substantial amounts of money, European grants were obtained by teams 2, 5, 7 only, and private company support was found by teams 2, 4, 7 almost exclusively.

### EVALUATION AREA 3: SCIENTIFIC PRODUCTION

#### Assessment on the scientific production of the unit

U900-CBIO scientific production is outstanding with more than 540 original articles published, 31.5% of which as corresponding authors. A number of papers were published in prestigious journals (Nat Genetics, Genome Biol, Nat Comm, Cell Rep, etc.). This impressive scientific production also includes well-respected specialized journals (Bioinformatics, Biostatistics), and authorship is well balanced among all the teams and researchers (junior and senior).

1/ The scientific production of the unit meets quality criteria.

2/ The unit's scientific production is proportionate to its research potential and properly shared out between its personnel.

3/ The scientific production of the unit complies with the principles of research integrity, ethics and open science. It complies with the directives applicable in this field.

## Strengths and possibilities linked to the context for the three references above

### *1/ The scientific production of the unit meets quality criteria*

Over the 2017-2022 period, U900-CBIO researchers were associated with 546 original articles, 31.5% of which as corresponding authors, sixteen reviews, fifteen books or book chapters, 45 software packages, one public database, two patents, and they delivered >300 oral presentations. Due to U900-CBIO positioning, this outstanding scientific production includes methodological papers (Montagud, eLife, 2022; Lazard, Cell Rep Med, 2022; Abecassis, Nat, Comm, 2021; Martinelli, Bioinformatics, 2021; Mozzgunov, Biostatistics, 2022) as well as important applications (Phelan, Nat Genet, 2017; Patel, Cancer Res, 2020; Molino, Nat Med, 2022; Jeannot, Clin Cancer Res, 2021). Moreover, generally accessible resources are released (Kondratova, Nat Comm, 2019). A large number of articles were published in high profile journals (Nature, Cell, Nat Med, Nat Genetics, eLife, Genome Res, Clin Cancer Res, etc.) or in well-respected specialized journals, which are very relevant for methodological developments (Bioinformatics, Biostatistics, Int J Biostat, PLoS Comput Biol, etc.). The large number of oral presentations (>300) further attests of the scientific trust by the community.

Due to the strong link with experimental technology developments that drive cancer research significantly, U900-CBIO scientific production is often at the forefront of its domain. The importance given to machine learning and now AI in the scientific orientation of the unit is very relevant and timely. Efforts to pave the way to the construction of digital twins in the future by team 2 are ground breaking.

### *2/ Scientific production is proportionate to the research potential of the unit and shared out between its personnel*

The unit is integrated within IC, which is a national hub in fundamental and clinical cancer research. That is, the potential and the opportunities are massive. U900-CBIO managed to capture this energy and also to contribute to its generation very efficiently. This is easily demonstrated by the impressive external funds that were raised (18M€, 90% of the unit's budget), the numerous collaboration papers including in excellent journals (Nature, Cell, etc.), the important own production (>125 last-authored papers), as well as the implication in large structuring efforts (IHU, RHU, PRAIRIE, ParisSanté Campus, SIRIC).

The different teams all raised important amounts of money and established vast networks of collaborators. They are hence in excellent positions to publish at a very competitive level. This has been clearly the case for the senior teams (1, 2, 4, 6, 7). The junior team 5 (Ballesta) has several publications as last author already. The junior team 3 (Cavalli) is at an earlier stage, but due to the extension of her ATIP grant, she should not be evaluated yet.

Overall, there is an excellent match between the potential offered by the IC integration and the actual unit scientific production. First and even last authorships seem balanced between the team members. PhD students and postdocs usually first author, and senior team members alternate as last corresponding authors.

### *3/ The scientific production of the unit complies with the principles of research integrity, ethics and open science*

IC, the hosting and supervising institution of U900-CBIO, has strict rules regarding scientific integrity and results reproducibility. In 2015 IC signed the French Charter of Ethics for Research Practices. March 2021 saw the creation of a dedicated board and the appointment of a scientific integrity officer. These rules are naturally implemented at U900-CBIO. IC and the doctoral schools offer specific training and seminars in scientific ethics, and all the PhD students to attend.

Moreover, U900-CBIO works with FAIR principles (Findability, Accessibility, Interoperability, and Reuse) regarding data. The latter are centralized in a system where each data set is affected to a clearly identified project and described with metadata. Tracking of developments and data analyses are achieved thanks to Git, GitLab, or Jupyter notebook. Most software is released in public repositories (e.g., GitHub), as open source, with appropriate licences (GNU-GPL, CeCILL, BSD3). Patient data are pseudo-anonymized (according to GDPR Regulation (EU) 2016/679) or anonymized whenever possible (using 95/46/CE directive and G29 opinion 05/2014 guidelines). Projects dealing with patient data are submitted to an internal review board (named CRIData), which checks the compliance of the project with respect to regulation laws, ethic rules, patient consents, etc.

So far, little wet lab activity is conducted on site (team 1 only), but soon teams 3 and 5 will start their labs. In every case, IC rules again apply or will apply since these laboratories are/will be located within IC.

## Weaknesses and risks linked to the context for the three references above

No real weak point. No specific, well-defined scientific objective that would be shared by several teams beyond technical shared necessities or the access to specific skills from another team.

## EVALUATION AREA 4: CONTRIBUTION OF RESEARCH ACTIVITIES TO SOCIETY

### Assessment on the inclusion of the unit's research in society

The U900-CBIO unit members are very active in connecting with the non-academic world, in particular many connections exist with the economic actors (pharma, biotechs), and the general public. This activity is outstanding.

- 1/ The unit stands out for the quality and the amount of its interactions with the non-academic world.*
- 2/ The unit develops products for the cultural, economic and social world.*
- 3/ The unit shares its knowledge with the general public and takes part in debates in society.*

## Strengths and possibilities linked to the context for the three references above

### *1/ The unit stands out by the quality of its non-academic interactions*

The unit has established excellent interactions with the economic world. A rather large number of CIFRE PhD fellowships (6) were obtained with different partners, large pharmas or small companies (SANOFI, Servier, SynSight, etc.). The CUBIC core platform was involved in the design of IC current compute server that was offered by Intel in exchange for learning better the needs of life sciences. A number of companies made donations, primarily to support PhD theses. Lastly, several teams are involved in clinical trials with companies, e.g., team 6 with MSD supporting the PEVOsq trial.

During the COVID-19 crisis, although not related to cancer, U900-CBIO team 2 contributed its expertise in assembling disease maps to address the new disease based on virus-human interactions (Ostaszewski, Mol Syst Biol, 2021).

### *2/ The unit develops products for the socio-economic world*

Interactions with the socio-economic world are excellent.

CUBIC developed a bioinformatics pipeline to process a gene panel sequencing data in partnership with Agilent that develops the latter panel under the name DRAGON. The CIFRE PhD fellowships also convey knowledge and expertise transfer towards economic actors.

Through its support to IC technology platforms and patient sample analysis, the unit has an impact on the hospital operations (prix Curie 2020 for the hospital group to team 6 Delphine Hequet). Similarly, the support to Curie scientific research that is targeted towards cancer, a major public health challenge, has long-term societal impact (prix Curie 2021 for the research center to CUBIC Nicolas Servant).

Along the same line, team 6 develops standards in cancer data meta-analyses and endpoints assessments (DATECAN initiative). Team 1 members participate in expert groups of the Genetic and Cancer Group (GGC, Unicancer) since 2017. They also participate in standardizing medical management practices for familial cancer syndromes designing gene panel test (Moretta, Bull Cancer, 2018; Dhooge, Eur J Med Genet, 2020). There is further ongoing work to review and evaluate the clinical utility of testing genes involved in kidney cancers and pancreatic cancer predisposition.

The many software packages (44) released open-source and the databases developed by U900-CBIO teams are freely available to every actor in science, including companies.

Several team leaders and permanent researchers teach at various institutions in Paris.

### *3/ The unit shares its knowledge with the general public and takes part in debates in society*

The unit communication towards the general public is excellent.

The unit leverages on IC and Mines Paris - PSL communication departments to target different audiences using social media, rewriting for non-specialists, and organizing interviews. Also, they regularly publish on their institutional websites notes or articles. Typical media used are scientific popularization media such as “Sciences et Avenir” and “La Recherche”, radio shows (France Culture “La méthode scientifique”). Papers were also published in general audience media such as Marie-Claire, Le Figaro, Le Monde, Paris Match, and Europe 1 and France Inter radio shows, and France 5 TV channel. U900-CBIO also participates in events like Fête de la Science, Paris Healthcare Week, or in Cité des Sciences et de l'Industrie.

Some specific examples are: Inna Kuperstein (team 2) published with Marina Thizeau a book to explain cancer to young children “Monsieur cancer veut toute la place”, Paris, Le Buveur d'Encre (coll. esprits curieux). A Ballesta (team 5) made an online video for the website of France culture, intended to popularize “Cancer chronotherapy” for the general audience (June 2018). A Ballesta also participated in the Radio program « La Méthode Scientifique » on France Culture (June 2018) and gave an interview for the CQFD Radio show by Stéphane Délétroz on the RTS Swiss Radio (Oct 2019) to popularize chronobiology and chronotherapies.

On September 29, 2022, for the 8th Scientific Summit held at the 77th United Nations General Assembly in New York, IC reminded that it is urgent to act against cancer. Examples of prevention strategies implemented in East Africa, Colombia and Serbia, and international clinical and translational research on cervical cancer were presented. C. Le Tourneau (team 6) organized and chaired a session at the United Nation entitled “Cervical cancer as a revealing example of worldwide disparities in cancer prevention and care”.

### Weaknesses and risks linked to the context for the three references above

Not really, but the limited number of patent applications filed. Are there granted patents?

## ANALYSIS OF THE UNIT'S TRAJECTORY

The U900 unit has been created in 2008 to provide expert support integrating and exploiting high-throughput data in cancer research and the clinic. Over the last mandate, the unit has been very successful by covering many essential aspects of data analysis in cancer, delivering state-of-the-art methods, and deploying them in applications with researchers and clinicians. This has for instance enabled U900-CBIO to be involved in many large institutional programs (CASSIOPEA, MYPROBE, epCART, Women's Cancer IHU). In line with the motivations that led to its creation, the unit has kept a strong methodological orientation meaning that the identification or development of the best method precedes its application, and excellence is sought through this strategy. This principle, which obviously makes a lot of sense, will remain at the heart of the unit strategy for the future. The inclusion of a core group (CUBIC, team 7) is an efficient strategy to streamline purely technical support and established methods, while preserving research teams time for deeper innovation.

The strong commitment to large-scale data management and analysis supported by CUBIC and its partnership with Intel will continue. The emergence of spatial biology and multimodal data acquisition has been identified as a priority for method development and cancer biology research support. Although this revolution is happening everywhere, it is nonetheless very timely. IC needs strong support and U900-CBIO is well positioned to make visible contributions despite the international competition.

The same is true regarding machine learning and especially AI. The unit has a long history and solid teams active in statistical method application and development. During the last mandate, AI has been integrated and C900-CBIO managed to position itself very well within the Paris scene obtaining four chairs in the PRAIRIE program. The integration of several teams with ParisSanté Campus digital health cluster will also provide many advantages and opportunities to collaborate.

There is a clear vision separating research-devoted technologies such as spatial biology and multimodal measurements from technologies that can be deployed on a large scale in clinical trials or to handle patients. Although this boundary could move with time, new cutting-edge technologies will emerge and the need to distinguish two types of methodological development, dealing with cohorts of very different sizes, will remain. U900-CBIO teams address the two types of needs and plan to continue doing so.

Before, only team 1 (Andrieu/Lesueur) had some wet lab activity on site to measure mutational profiles. A major change in the unit's strategy is that the two junior teams (Ballesta and Cavalli) will have both wet and dry labs as soon as the new space at Saint-Cloud is available (2025). We believe that it is an excellent idea and it will generate a new type of project. It is also very much in line with the fact that more and more younger biologists have also acquired pretty solid training in data analysis or even hard sciences in some instances. To give such profiles the opportunity to deploy their full potential in France is essential to keep up with international competition.

Medical informatics is booming and well supported in Paris, e.g., through ParisSanté Campus. U900-CBIO strategy for the future does not mention it. It would be useful to know its position, future potential integration in a new junior group for instance, or the decision to only collaborate on this.

To finish, from an organizational perspective, the multi-location setting of U900-CBIO is not ideal. Many efforts were made and will continue to be made to integrate its people and given them a sense of belonging to a same group. Computational units with strong methodological expertise but devoted to cancer research can be found at several high profile places internationally. They are not common in France unfortunately, and IC/Mines/Inserm should try to alleviate as much as possible the burden caused by multiple locations. The situation with administrative staff (only two people currently with one to retire very soon and no immediate replacement) is not adequate for such an active unit. The same could be said regarding the recruitment of a few software engineers and informaticians to maintain software packages and the infrastructure more efficiently.

## RECOMMENDATIONS TO THE UNIT

### Recommendations regarding the Evaluation Area 1: Profile, Resources and Organisation of the Unit

In terms of profile, a nice addition that could further potentiate data access and opportunities for patient-related studies, would be to recruit a medical informatics team leader. A junior professor chair by Inserm might be a good tool for such an operation.

In terms of resources, the obvious weakness is on the admin personal. Two people for a unit of the size of U900-CBIO, with several EU grants that necessitate additional work compared to national funding, is not enough. Even worse, one such person has already announced her retirement and no immediate replacement is planned. We strongly recommend solving this issue that will ultimately penalize a very productive group of scientists. Along the same line, engineers responsible to maintain software developed by the unit and who would hold a permanent position would catalyze U900-CBIO visibility. It would stabilize knowledge and know-how instead of relying on PhD student or postdocs whose ultimate fate is to leave the unit. Lastly, the unit should search for lab technicians (engineers) holding IC positions to support the future wet lab activities of the junior teams.

Despite the cost, more frequent unit-wise retreats should be considered to alleviate the multi-site inconvenience and foster a sense of unity.

### Recommendations regarding the Evaluation Area 2: Attractiveness

The unit is already very attractive. Towards young parents in science, a fund could be constituted to offer nanny services while attending scientific meetings or giving lectures.

### Recommendations regarding Evaluation Area 3: Scientific Production

The overall production is already excellent, but certain teams, e.g., teams 3 and 5, should now aim at increasing their numbers of publications in high profile journals.

Team 7 should generate more papers to increase visibility on the many tools and pipelines they develop. There are dedicated journals for technical contributions in bioinformatics (PLoS Comput Biol, BMC Bioinformatics, Ddatabases, etc.) that should welcome such submissions. To publish a pipeline after the princeps research paper that motivated its development should provide data to exemplify the pipeline application. A better effort giving visibility to more team 1 members as co-authors should be considered.

### Recommendations regarding Evaluation Area 4: Contribution of Research Activities to Society

In this area, the unit is already quite active through different media and actions. We have no specific recommendation but to continue the effort.

## TEAM-BY-TEAM OR THEME ASSESSMENT

**Team 7:** Institut Curie Bioinformatics Core Facility (CUBIC)  
Mr Emmanuel Barillot

### THEMES OF THE TEAM

The CUBIC core facility develops multidisciplinary expertise in bioinformatics and biostatistics to support Institut Curie biotech core facilities (Curie Core Tech), research units and hospital teams. They are involved in data analysis, data management, data interoperability for omics and biomedical data, quality management and best practices, and support for new technologies. They are also involved in support to HPC and coordination of the single cell activities and the Bioinformatics Hub at Institut Curie.

### CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

The CUBIC team (core facility) was not directly evaluated in the previous report.

Relevant comments from unit evaluation:

*The different duties of the bioinformatics platform within the unit (service to other IC units, support to the unit's research teams, own research and development) should be clarified internally and evaluated (as a unit's platform) in order to evidence its central role.*

Indeed, the CUBIC facility is now considered one of the evaluated teams and has reported on their activities like other teams.

*The R&D and internal support missions of the bioinformatics platform should be increased in order to prepare future accurate analysis pipelines for other IC units.*

Many pipelines have been developed to support data analyses in the context of routine activities as well as research projects. These pipelines have been developed following best practices and reproducibility. CUBIC is involved in supporting new technologies to help other teams, and coordinating bioinformatics activities at Institut Curie.

### WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	0
Maîtres de conférences et assimilés	0
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	0
Personnels d'appui à la recherche	17
<b>Sous-total personnels permanents en activité</b>	<b>17</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	0
Personnels d'appui non permanents	0
Post-doctorants	0
Doctorants	0
<b>Sous-total personnels non permanents en activité</b>	<b>0</b>
<b>Total personnels</b>	<b>17</b>



## EVALUATION

### Overall assessment of the team

The scientific production of the team is excellent to outstanding, with around 80 co-authored publications in journals and conferences over 2017-2022 and a number of publicly available software packages, several of them widely adopted.

The attractiveness/impact is excellent with involvement in a dozen grants, a dozen invited talks, twenty hosted interns and two co-supervised PhD students, the "prix Curie" awarded to N. Servant in 2021, involvement in many workgroups and definition of best practices for pipeline development, a few outreach activities.

The facility is recognized for its technological expertise, involvement in training, and support to other teams/facilities and coordination of bioinformatics activities at Curie.

### Strengths and possibilities linked to the context

The team includes nineteen permanent staff members.

The team has a high level of scientific productivity reflected in 80 co-authored publications in a spectrum of journals. The team develops a large number of software packages, most of them regularly updated. Usage statistics show that several of these are widely adopted internally and/or externally. The team is involved in defining best practices for the development of bioinformatics pipelines. The facility obtained COFRAC certification in 2017 and ISO9001/NFX50-900 certification in 2022 for clinical diagnostic & HTS data analysis.

The team was associated to a dozen grants as task leader across 2017-2022.

Team members were invited to give a dozen seminars/talks across the period. One of the current operational managers received the internal "prix Curie" in 2021. Team members have grant reviewing activities for several agencies and some reviewing activities for journals. The team co-organized five conferences/workshops over the period.

Team members are strongly involved in training, on average around 200 hours per year, two thirds external. The facility is a member of IFB and a partner in MUDIS4LS, and a member of international alliance Core4Life.

Over the assessment period, the team hosted 2 co-supervised PhD students and 20 interns.

### Weaknesses and risks linked to the context

The small number of publications where lead authors belong to the core facility seems surprising, given the large number of software developments.

It seems that a large part of the successes of the team rest on the shoulders of one single member, Nicolas Servant, who appears to be a co-author in a majority of publications, together with two or three other members, in a lesser extent. He also gave most of the invited talks. This situation raises the question, either of an unbalanced distribution of the work within the team, or a less recognition of the work of others.

The facility handles a wide range of diverse projects, with constant evolution of the techniques that required adapting to new techniques (new data types) and increasing data volumes.

The generalized shortage of skilled bioinformatics experts leads to difficulties in recruitment and stability of the team.

The team reports the risk associated with funding permanent positions with grants.

o Ask for details about this / the business model for this core facility

There is no information about the proportion of women in the team, and concrete measures to correct for a possible imbalance, as it is often the case in the field.

Career perspectives for engineers are not mentioned and could impact the attractiveness of the proposed positions.

It is not mentioned if engineers are subject to a (regular) scientific evaluation of their activity, and if yes, which criteria this evaluation relies on.

## Analysis of the team's trajectory

The team's trajectory aims to tackle the evolution of omics technologies with new data types, increasing need for data integration, increasing volumes and diversity of data. Five directions are presented.

- o Data management: archiving, compliance with health data warehouses
- o Continued quality management and best practices (including continuous integration)
- o Supporting emerging technologies (spatial omics and multiplex image analysis, data integration, Nanopore long read sequencing, single-cell omics)
- o Artificial intelligence (multi-omics data integration, internal collaborations)
- o Coordination of bioinformatics at Curie (promotion of the Bioinformatics Hub)

The team's trajectory seems logical and ambitious with respect to their expertise and the current technological advances in life sciences.

## RECOMMENDATIONS TO THE TEAM

The business model and publication strategy of the platform should be clarified. It is also unclear whether they plan for more openness to teams outside Curie (beyond public software development).

The sustainability of positions (funding, attractiveness) and possible plans for extending the facility should be closely monitored.

The team has been evaluated according to criteria adapted to the evaluation of research entities. Some specific information is lacking, such as the number of projects submitted to the platform, decision process, project management process, user satisfaction.

**Team 1:** Genetic Epidemiology of Cancers (GEC)  
Ms Fabienne Lesueur

## THEMES OF THE TEAM

The GEC team is focusing on the characterization of genetic variants involved in a familial cancer mainly breast and ovarian cancers but also thyroid cancer by considering effects of other genetic and non-genetic (lifestyle and environmental exposures) factors. It includes the following topics: (1) cancer risk of BRCA1/2 pathogenic variant carriers and their modifying factors, 2) assessment of clinical utility of breast cancer susceptibility genes other than BRCA1/2 testing, 3) identification of new cancer susceptibility alleles in familial breast cancer with no BRCA1/2 pathogenic variant and b- differentiated thyroid cancer, 4) ATM and cancer risk. For that purpose, GEC team initiated and coordinated six national studies.

## CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

Team 1 responded well to the first recommendation of defining new research niches by developing their research on the role of ATM in cancer and new research axis on thyroid cancer. They pursued the development of the PIGE ((platform of investigation in genetics and epidemiology) database and organized the first ATM and cancer risk international workshop to increase its visibility.

The GEC team also responded appropriately to the second recommendations of 1) increasing scientific staff that can write projects and papers by recruiting 2 post-docs and hosting 3 visiting scientists over the evaluated time period; and 2) maintaining the consistency within the group by organizing weekly team meeting.

The team responded appropriately to the third recommendation (that was in part related to or redundant with the first recommendation on research niches and PIGE database development) by strengthening their collaborations with other teams of the CBIO unit (7 articles in collaborations).

## WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	0
Maîtres de conférences et assimilés	0
Directeurs de recherche et assimilés	1
Chargés de recherche et assimilés	1
Personnels d'appui à la recherche	8
<b>Sous-total personnels permanents en activité</b>	<b>10</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	0
Personnels d'appui non permanents	0
Post-doctorants	0
Doctorants	2
<b>Sous-total personnels non permanents en activité</b>	<b>2</b>
<b>Total personnels</b>	<b>12</b>

## EVALUATION

### Overall assessment of the team

The team consists of two groups: a research team and an investigation group devoted to PIGE database and data collection. During the period, the team supervised the formation of five PhD students and hosted two post-docs and three visiting scientists. The team coordinates six major national collections on breast and ovarian cancer, ATM and the PIGE database. The scientific production is excellent with a good international visibility, and an excellent capacity of raising national funding. The team is well integrated with the social, economic and clinical environment which favor the transfer of knowledge from research to clinical interventions.

### Strengths and possibilities linked to the context

The research team's activities are organized among three main themes: breast cancer with and without BRCA1/2 pathogenic variants, ATM and cancer risk, and thyroid cancer. The team is composed of twelve people structured in two well-separated groups: one dedicated to the collection of data (6 engineers or assistants) and a second on the research activity (2 Inserm researchers, 3 engineers with permanent position and 2 PhD students).

The team benefits from the proximity of genetic department at Institut Curie, clinicians, biologists as well as pathologists due to its location during nearby Institut Curie during the past five years. It allowed the initiation and coordination of six national studies, the development of the PIGE database as well their strong contribution to international consortium: co-leading one international consortium (thyroid cancer) and belonging to several other international consortia (breast and ovarian cancer) where they have strong involvement in scientific projects

The team benefits from various financial sources (11 grants among which 8 as coordinator) including four ongoing national grants (2 INCa, 1 Ligue Nationale contre le Cancer et 1 Fondation ARC) to pursue their research.

The team has an excellent scientific production regarding the size of the team (2 senior researchers) with 26 articles with leading position among a total of 57 articles

Over the evaluated period, the team hosted 2 postdoctoral fellows, three foreign visiting scientists (from USA, Greece and Brazil) and trained 5 PhD.

### Weaknesses and risks linked to the context

There is a crucial need to recruit, attract new and young scientists with permanent position. Indeed, the team includes only two scientists with permanent position among which one is going to retired during the next contract.

A small staff of six people (engineers or assistants) including two with non-permanent position, to maintain and develop the collection of the six national studies and the PIGE database.

The number of publications of PhD student as first author is low, two PhD students among the three who have defended their thesis, having a single paper as first author.

### Analysis of the team's trajectory

The GEC team will pursue its research projects to characterize genetic and non-genetic factors involved in cancer initiation. Their specific aims will be: 1) to evaluate the potential clinical utility of genes for HBOC syndrome, 2) to assess the performance of polygenic risk scores (PRS), 3) to improve cancer risk assessment in women predisposed to HBOC syndrome by studying interactions between genetic risk factors and other risk factor, 4) to characterize ATM-deficient tumors.

The five-years project is ambitious given the human resources and requires the recruitment of new personnel, which is anticipated with two open postdoctoral positions and one open position for a junior professor in genetic statistics/genetic epidemiology.

## RECOMMENDATIONS TO THE TEAM

The scientific production of the team and its role in breast cancer risk research should be maintained. The development of new research field on thyroid cancer should be pursued.

The training of PhD students should be continued at least at the same level and their participation to publications, potentially through partnerships within CBIO unit, could be increased.

The recruitment of new scientists with permanent position is necessary regarding the retirement of one of the senior scientists during the next contract. The recruitment of postdoctoral fellows needs to be maintained and possibly increased.

The collaborations with others teams within CBIO unit could be increased.

The level of funding acquired over the past five years needs to be maintained and potentially increased with to ensure the achievement of the projects and the maintain of the national studies and PIGE effort.

The storage of the sample/biological collections and of the massive amount of data to be generated needs to be ensured especially with the move of the team to another site.

**Team 2:** Computational Systems Biology of Cancer (SYSBIO)  
Name of the supervisor: Mr Emmanuel Barillot

## THEMES OF THE TEAM

The SYSBIO team focuses on computational approaches to the systems biology of cancer. Four complementary research lines in computational cancer systems biology are considered specifically:

1. The formalization of cancer knowledge towards mathematical models.
2. The development of simulation Boolean/logical models of cancer states and transitions that incorporate knowledge of cancer mechanisms, which includes processes such as tumor onset and progression together with models of the spatial organization of tumors.
3. The integration through machine learning of complex multiomics data of tumors (from sequencing to images) across different scales.
4. The development of software tools to support the previous lines.

These approaches are then applied primarily to different types of carcinomas and sarcomas, although some other application areas are considered as well.

## CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

A – Recommendations on scientific production and activities (criterion 1)

The team should continue to target highly visible journals, explore more opportunities for non-academic partnerships and increase the scientific productivity of PhD students.

The productivity of the team increased further by about 20% over the past five years. The productivity of the PhD students is strong with on average three articles per PhD. Some partnerships have been established with non-academic partners, specifically two pharmaceutical companies.

B – Recommendations on the team's organization and life (criterion 2)

The team would need to attract a permanent full-time researcher to support the proposed project.

An additional team member has been recruited on a permanent basis two years ago and his funding is supporting through a sequence of contracts.

Promotion of women staff should be encouraged.

Dr. Calzone has been advanced to Joint PI of the group, which is a substantial gender balancing step.

C – Recommendations on scientific strategy and projects (criterion 3)

Further development of the ACSN resource using text-mining methodologies would need a quite specific expertise that may require an external collaboration.

This recommendation has not been heeded. Given the multiple possible approaches to the integration of complex omics data, some strategic choices need to be made to avoid excessive dispersion of the team.

Collaboration with a biostatistics team would be a plus for the project line on the mathematical modelling of cell fate.

Some tangible efforts have been made to integrate the data into the mathematical models of tumor states and transition developed by the team. This remains an ongoing effort as such development are still largely an open problem.

The team should better prioritize its software development activity in order to secure the long-term upgrading and maintenance of the most strategic tools. This may imply to define a clear policy on stopping maintenance for some of the software.

Important efforts have been carried out to ensure the long-term maintenance of software tools along bioinformatics best practices. Open-source practices have been used as a key strategy, which ensures easy accessibility of the tools. Interoperability is strengthened through the use of open standards and reusability is increased through the packaging of test units and documentation. The availability of a permanent staff scientist with development skills is a plus, although the availability of a dedicated software engineer remains desirable.

## WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	0
Maîtres de conférences et assimilés	0
Directeurs de recherche et assimilés	1
Chargés de recherche et assimilés	0
Personnels d'appui à la recherche	6
<b>Sous-total personnels permanents en activité</b>	<b>7</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	0
Personnels d'appui non permanents	0
Post-doctorants	3
Doctorants	8
<b>Sous-total personnels non permanents en activité</b>	<b>11</b>
<b>Total personnels</b>	<b>18</b>

## EVALUATION

### Overall assessment of the team

- Global: Excellent to outstanding
- Strong international niche in mathematical modeling of cancer regulation
- Outstanding international embedding
- Outstanding funding acquisition
- Excellent/outstanding attractivity given outstanding national/international visibility and robust funding

### Strengths and possibilities linked to the context

The team presents a solid backbone with one director of research, two engineers of research being considered as searchers, three support engineers of research (two of them being paid on our grants), three postdoctoral researchers, and six PhD students. It is highly productive with over 90 articles published over the period, half of which (co-)led by the team. This also represents a significant increase over the previous period. The team is well positioned at an international nexus in computational biology of cancer with links to strong international partners. The participation to for example the Exascale Centre of Excellence in Personalised Medicine PerMedCoE indicates a strong international profile. This is also reflected in the strong acquisition of very competitive European funding via multiple large projects. The strong occupancy of the niche of Boolean/logical modeling of oncogenic processes is a solid asset in a highly competitive area such as computational biology of cancer where many high-profile teams are present internationally. It can provide a key element to sustain an attractive position on the international scene and an expertise that can be uniquely leveraged in large international projects.

The team has strong local, national, and international visibility. Two PRAIRIE chairs are held by members of the team. Teaching covers multiple local institutions, which provides a much-needed conduit for recruitment. Multiple international workshops have been organized, which gives the team further international visibility. Almost one hundred oral presentations were given, over a third of them invited. Multiple pieces of software are developed and supported, several of them with strong international visibility.

The deployment of the expertise in the formalization of biological knowledge (originally for cancer) towards COVID during the pandemic has been a significant societal contribution. Further contributions relevant to society include multiple interviews and a children book on cancer mechanisms.

## Weaknesses and risks linked to the context

It appears that guaranteeing good human resource continuity is challenging, not so much at the level of funding, but rather at the level of recruitment. The strong rebalance between postdoctoral level and doctoral level staff over the current period is a risk for the integrity of the expertise of the team. While both models are sustainable, long-term stability is important. Although a broader teaching portfolio contributes to improved recruiting of doctoral students, recruiting at the postdoctoral and engineer level remains a bottleneck.

While collaborations across Institut Curie provides multiple opportunities to access exciting data, in particular relevant to omics data integration, Boolean/logical modeling requires experimental designs different from those typically used by biological teams. Lack of expertise and resources in running biological experiments and collecting data is a risk. It is also potentially an important opportunity as it could make the work carried out by the team truly unique.

## Analysis of the team's trajectory

The proposed trajectory builds in a coherent fashion further on the expertise and strengths of the team. In the first axis, further development of multimodal data analysis approaches is a natural progression, in particular further extensions to multilinear approaches (i.e., tensor factorization) and nonlinear approaches (i.e., deep learning methods). It would however be necessary to develop a specific profile in this crowded niche. The project on immunotherapy provides a good opportunity to strengthen the biological relevance of the methodology. For the second axis, the Boolean/logical modeling of the tumor microenvironment / cellular ecosystem could be a major opportunity. There is strong interest in improving the understanding of these complex processes. There are only limited conceptual and computational tools to bring together all data and knowledge over multiple scales, which is required if one is to achieve relevant understanding of these processes. As such, a well-structured methodology to tackle these problems could find broad interest across the whole cancer community. The third axis, with the development of digital twins, is still very much a stretch goal that could carry the team forward beyond the next.

The development of patient-specific Boolean/logical models can provide a niche where strong international leadership is possible. It also provides a framework that can binds the model-driven and the data-driven approaches developed by the team, connects with several other teams of the unit (for example towards the modeling of drug response in cancer treatment), as well as across the Curie Institute.

## RECOMMENDATIONS TO THE TEAM

- \* Bridge further towards other teams of the unit. Cross-leverage expertise between teams.
- \* Further build out the profile on Boolean/logical mathematical models of cancer processes. Explore further synergies between modeling approaches and data-driven approaches.
- \* Sustain strong international position by strengthening a unique profile and maintaining strong collaborations.



**Team 3:** Computational Biology and Integrative Genomics of Cancer (CBIGC)

Name of the supervisor: Ms Florence Cavalli

No evaluation due to team extension as ATIP Avenir

## THEMES OF THE TEAM

## CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

### WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	0
Maîtres de conférences et assimilés	0
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	1
Personnels d'appui à la recherche	1
<b>Sous-total personnels permanents en activité</b>	<b>2</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	0
Personnels d'appui non permanents	0
Post-doctorants	1
Doctorants	2
<b>Sous-total personnels non permanents en activité</b>	<b>3</b>
<b>Total personnels</b>	<b>5</b>

## EVALUATION

Overall assessment of the team

p\_Overall assessment-T4

Strengths and possibilities linked to the context

Weaknesses and risks linked to the context

Analysis of the team's trajectory

## RECOMMENDATIONS TO THE TEAM

**Team 4:** Statistical Machine Learning and Modeling of Biology Systems (CBIO)  
 Name of the supervisor: Mr Thomas Walter

## THEMES OF THE TEAM

The team develops Machine Learning and Computer Vision methods for biological and medical applications. The applications can be categorized into 3 groups: fundamental biology (analysis of FISH images, of genomics data to identify horizontal gene transfers), drug design/discovery (chemogenomics, protein-ligand interaction screening by 3D docking of candidate drugs, quantifying complex cellular phenotypes in drug screening, systems biology) and precision medicine (identification of molecular biomarkers, predictive models for diagnosis, prognosis and treatment response prediction). The team is involved in many collaborations with biologists and medical teams.

## CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

### A – Recommendations on scientific production and activities (criterion 1)

The top quality of the scientific production of team 2 would benefit from a better strategy in terms of software development. The diversity of the produced code could be reduced, and software diffusion and maintenance improved.

The diversity of the produced code is linked to the diversity of applications that the team is working on, and therefore has not been reduced. However, the team has worked on better software diffusion and maintenance e.g. with the development of FISHquant v2 for analysis of FISH data.

### B – Recommendations on the team's organization and life (criterion 2)

Team 3 should continue to strengthen its collaboration with other teams in the lab.

More collaborations have been established in the last 5 years with other U900 teams (Barillot team, CUBIC core facility). 2 co-supervised PhD students in the last 5 years.

It would be beneficial to expand the team with another member (preferably senior), allowing a better distribution of the supervision load.

Although the former team director left in 2017 to join a private company, the team hired a junior PI in 2021. Two more positions should have been or soon be filled (junior PI in 2023, assistant professor in 2024).

Ask the team about these positions.

It would also be of interest to have more space for the whole lab while avoiding multi-localization.

Only foreseen for 2028 when the team should move to the ParisSanté Campus (Val de Grâce).

### C – Recommendations on scientific strategy and projects (criterion 3)

The team's science strategy is strong relative to its past achievements and today's emerging technologies (2017). Benefiting from privileged access to data, research in machine learning applied to medicine and systemic biology is certainly a worthwhile project. Nevertheless, more theoretical research should not be left behind.

Especially since the former team director left, the team strategy has evolved towards more applications and fewer theoretical developments, the latter being exclusively directed towards biological/medical applications. The team has obtained two PRAIRIE (artificial intelligence) chairs. The team has one patent (HRD) and another patent application in December 2022.

## WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	2
Maîtres de conférences et assimilés	1
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	1
Personnels d'appui à la recherche	1
<b>Sous-total personnels permanents en activité</b>	<b>5</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	2
Personnels d'appui non permanents	0

Post-doctorants	1
Doctorants	10
<b>Sous-total personnels non permanents en activité</b>	<b>13</b>
<b>Total personnels</b>	<b>18</b>

## EVALUATION

### Overall assessment of the team

The scientific production of the team is excellent, with 90 publications in journals and conferences over 2017-2022.

The attractiveness is excellent to outstanding with cumulated funding of 3.6 million euros overlapping with 2017-2022, over 100 invited talks, sixteen PhD students trained to completion, two PRAIRIE chairs and one young AI woman engineer award.

The societal impact is excellent with one patent application, several partnerships with the private sector (joint PhD students and contracts) and a number of outreach activities.

The team is recognized for its dynamism in methodological research, collaborations with biologists and medical doctors, and teaching.

### Strengths and possibilities linked to the context

The team includes four permanent researchers, ten PhD students and three postdocs/non-tenured researchers.

The team has a high level of scientific productivity reflected in 90 publications published across a spectrum of journals and conferences. The team develops software, including two R packages, one Bioconductor package and the FISHquant V2 software available on Github.

The team consistently benefits from various financial sources, including two PRAIRIE chairs, several ANR grants, several private contracts (Cairn Biosciences, Janssen + CIFRE with Iktos, Sanofi) and funding from foundations, especially in the context of the application of their research to cystic fibrosis (FRM, Vaincre la Mucoviscidose, Fondation Dassault Systèmes, Fondation MSD Avenir).

Team members were invited to give more than 100 seminars or oral talks to international meetings. One team member received the Young AI Woman engineer award and another was named Officier des Palmes Académiques. Team members have grant reviewing activities for several agencies and reviewing activities for a number of journals and international conferences. The team co-organized five workshops over the period.

Team members are strongly involved in teaching, especially at Mines Paris. This enables them to recruit excellent students.

Over the assessment period, the team successfully trained seventeen PhD students to completion (4 of them co-supervised) and hosted five postdoc researchers. Two of the team permanent researchers defended their HDR in the period.

The team deposited one patent application in 2021.

### Weaknesses and risks linked to the context

The team's research covers a range of very diverse applications leading to a risk of overdispersion (projects and external requests).

Their peculiar administrative situation and diverse funding sources create admin overheads.

The team does not generate any data and depends on collaborations for external data generation.

The permanent staff dedicates a significant fraction of their time to teaching duties and take on significant PhD supervision commitments (10 PhD students for 4 permanent staff).

Artificial intelligence is a fast-moving field with strong competition (including from the private sector).

## Analysis of the team's trajectory

The team's trajectory aims to take advantage of their machine learning and AI expertise as well as the current context of new biotechnological developments. Three directions are presented.

- Translational research: 1/ continued collaborations with clinicians and pharma, 2/ work on method robustness and reproducibility (interpretability, uncertainty assessment).
- Novel methods for novel technologies. 1/ New data types: RNA-seq, electronic health records, TE variability & methylation (ANR project), highly multiplexed images, new microscopy (especially label-free). 2/ Scaling up methods for GWAS, post-selection inference; drug-protein interactions, chemogenomics; computational pathology; from bacterial genomics to metagenomics.
- Integrating complementary data types to combine information across scales. 1/ integrate systems biology and chemogenomics (drug targets and drug-protein interactions); apply to cystic fibrosis (funded) and triple negative breast cancer (collab IGR). 2/ DL for multimodality (NN to integrate multiple sources of information). Apply to: transformer for patient trajectories (medical data), biological network information to constrain NN architecture for omics data; in silico labeling to predict complex modalities; methods to query pathology data wrt cell phenotypes; integrate radiomics and multi-omics for treatment prediction.

The team's trajectory seems logical with respect to their expertise and the current technological advances in life sciences, but also very ambitious, considering the current diversity of their applications and their demanding teaching duties. The new directions are even more diverse. The ability of the team to achieve them might depend heavily on the planned recruitments.

## RECOMMENDATIONS TO THE TEAM

The focus on deep learning and neural networks seems rather new (compared to machine learning) and although it makes sense given increase in data size and complementarity, the team should also assess how to preserve their historical specific expertise (strategic positioning) and avoid overdispersion.

The integration of the newly recruited researchers should be monitored with care given the small size of the team, their broad thematic engagement and their multi-localization.

**Team 5:** Cancer Systems Pharmacology (SYSPHARMA)

Name of the supervisor: Annabelle Ballesta

## THEMES OF THE TEAM

The team develops complex mechanistic models to optimize and individualize treatments. They use mathematical and statistical methods to analyze multiscale and multitype datasets in order to build digital representations of organs, patients or populations, with the objective to optimize or individualize treatments. They focus on quantitative system pharmacology to dynamically model the pharmacokinetics or pharmacodynamics (PKPD) of a drug, using ordinary differential equations, statistical tools and machine learning technics.

They in particular focus on the individualization of combination of drugs using tumor multiomics and drug screening data, with application to the glioblastoma, and to the design of sex specific and personalized cancer chronotherapy according to the patient's circadian rhythm.

## CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

Not applicable - The team was created as a junior team during the evaluated period (Award Atip-Avenir 2018) and joined CBIO in 2021.

## WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	0
Maîtres de conférences et assimilés	1
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	1
Personnels d'appui à la recherche	0
<b>Sous-total personnels permanents en activité</b>	<b>2</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	0
Personnels d'appui non permanents	0
Post-doctorants	2
Doctorants	2
<b>Sous-total personnels non permanents en activité</b>	<b>4</b>
<b>Total personnels</b>	<b>6</b>

## EVALUATION

### Overall assessment of the team

The team has an excellent publications list (15 papers and 3 reviews), taking into account its very small size (only two permanent members), including nine as first or last author, and also one patent.

It has an excellent involvement in outreach, thanks to 10 interviews in various media, both professional and large audience.

It has an excellent PhD training, with two PhD defended during the period (for only two junior permanent members), two one going PhDs and various L3, M1 or M2 internships.

Funding is also excellent, with two large and competitive fundings (ATIP-Avenir and MIC ITMO), reaching 1 M€ as principal investigators, and more than 1M€ in shared fundings.

### Strengths and possibilities linked to the context

Taking into account the very small size, excellent in publication, outreach, PhD training, funding.

### Weaknesses and risks linked to the context

The main weakness of the team is its very small size. There is only two permanent members, who have "junior positions" (including one assistant professor who has teaching duties). The workforce is thus limited as well as the capacity to supervise PhD students. Small teams are very sensitive to random events and have a limited capability to respond to new opportunities.

The smallness of the team also leads to an overload of work on the permanent members to prepare and submit subject, write report.

Moreover, using machine learning techniques ("AI") is a very trendy and rapidly increasing subject. Many teams are being created and will work on these themes in the forthcoming years, including teams of large private companies. There is thus a risk for the team to be taken by speed by larger teams on these subjects.

Moreover, the team depends on external collaborations to obtain data, which is always a weakness, even if some of the collaborations are long standing and firm.

### Analysis of the team's trajectory

Not applicable - The team was created as a junior team during the evaluated period (Award Atip-Avenir 2018) and joined CBIO in 2021.

## RECOMMENDATIONS TO THE TEAM

The team's scientific production should be maintained, and the level of funding acquired over the last five years (other than Postdoctoral and PhD fellowships) would probably need to be increased to ensure that the team's projects can be carried out.

The recruitment of a new scientist with a permanent position (potentially with HDR) is necessary for the development of the team since up to now this team includes only two researchers with permanent position.

The training of PhD students should be continued at least at the same level and their participation to publications, potentially through partnerships within U900-unit, could be increased. It will be a good way to participate to their integration to the lab as the team is located outside the main core of the unit.

**Team 6:** STATistical Methods for Precision Medicine (STAMPM)

Name of the supervisor: Mr Aurélien Latouche

## THEMES OF THE TEAM

The team has two main axes. Axis 1 is the development of advanced statistical method related to the use of biomarkers, either in early phases trials or as prognostic or predictive markers

Axis 2 is the evaluation of medical practice based either on trials ore routine practice with IPD meta-analyses, evaluation of statistical practices and development or evaluation of new outcomes for new treatment evaluations with an highlight on immunotherapy. The team is involved in numerous collaborations with clinical teams.

## CONSIDERATION OF THE RECOMMENDATIONS OF THE PREVIOUS REPORT

A – Recommendations on scientific production and activities (criterion 1)

Numbers of publications in methodological/statistical journals, presentations at statistical conferences and release of software should gradually increase over time to improve the visibility of the team and national and international recognition.

Team has published in high level statistical journals, and developed collaborations with internationally recognized experts in the field.

More PHD students, and if possible post-doctoral researchers, should be hired in the future. They should be encouraged to publish their results in highly ranked journals, including more methodological publications, as well as to develop packages/software for use by the scientific community and dissemination of the results.

A total of ten PhD has been defended or are ongoing and several R packages have been developed (although not by PhD students)

B – Recommendations on the team's organization and life (criterion 2)

The team needs to progressively find its way of working its own organisation and life, benefiting from the interactions with the other teams, but at the same time becoming a team on its own. All possible efforts should be made by the team to obtain an adequate physical location in order to foster collaborations with other teams. The team would certainly benefit from attracting an additional permanent scientist in order to better guarantee the durability of the team. One1 professor (biostatistics) and two assistant professors (1 in biostatistics and 1 in surgical oncology) have been recruited, whereas 2 medical doctors with clinical activity have left the unit.

C – Recommendations on scientific strategy and projects (criterion 3)

The scientific projects are interesting and fit well in the framework of precision medicine. They should however be better structured and prioritized. In the future, it will be interesting to further develop collaborations with other teams, and, in particular, with team 1. Indeed, real synergies could certainly emerge from the complementary expertise of the two teams, on the methodological questions raised by the cohort database collected by team 1, and there is a clear potential for further exploitation of these data.

On of the researcher (Pr Le Tourneau) created a department of drug development and innovation to stimulate collaborations with clinical teams. Collaboration with the epidemiology team has been postponed until the team has sufficiently grown.

## WORKFORCE OF THE TEAM: in physical persons at 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	3
Maîtres de conférences et assimilés	2
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	0
Personnels d'appui à la recherche	0
<b>Sous-total personnels permanents en activité</b>	<b>5</b>
Enseignants-chercheurs et chercheurs non permanents et assimilés	1
Personnels d'appui non permanents	0

Post-doctorants	0
Doctorants	6
<b>Sous-total personnels non permanents en activité</b>	<b>7</b>
<b>Total personnels</b>	<b>12</b>

## EVALUATION

### Overall assessment of the team

Initially a junior team, the team has gained a sufficient size (5 permanent positions) and exhibit a very good publication track with 148 publications (though not all directly related to their research activity). It has national and international recognition in the field with strong collaborations and organization of national congress. Several fundings have been obtained and ten PhD defended/ongoing since the past contract. Moreover, several members of the team have and an important implication in teaching.

Finally, research axes are well aligned with the medical environment of the Curie institute and the current development of new therapies.

### Strengths and possibilities linked to the context

The team has gained a more stable size with five permanent positions. The team has a well-recognized expertise in the field of biostatistics both at a French and international level.

Axes developments are well aligned both with the medical thematic of Curie institute and the current development of immunotherapies and targeted therapies.

This ensures access to either public or private (CIFRE PhD) fundings.

### Weaknesses and risks linked to the context

Departure of two members of the team will implies some modification of axes development which require clarifications.

The team has to ensure an adequate balance between methodological support for clinical research project and its own research axes. This is critical since all members of the team have either clinical activities or important teaching assignment.

### Analysis of the team's trajectory

The team's trajectory is to get involved in a new axis of research linked to platform trials. This line of research is linked to developments made during the previous contract on biomarker x treatment interaction detection methods, which are not very applicable within the framework of classic trials because they require a high number of subjects.

Perspective:

- Platform trials on small sample size linker to a RHU on CART-cells for solid tumors;
- Development of methods for radiomic biomarkers;

These evolutions are logical extensions of the research areas already developed by the team for the former and a more new development for the latter. Nevertheless, it is not very clear which of the actual axes will be not continued, since this is a rather small team.

## RECOMMENDATIONS TO THE TEAM

Since the size of the team is still somehow small (5 permanent members but none with a 100% research activity) it could benefit from hiring new post-docs.

It would also benefit from a clarification on the evolution of the research project which is ambitious.



## INTERVIEW SCHEDULE

### CONDUCT OF THE INTERVIEWS

#### Date

**Start:** 15 décembre 2023 à 09h00

**End:** 15 décembre 2023 à 17h00

**Interview conducted: on-site**

## INTERVIEW SCHEDULE

### 15 Décembre 2023

- 9:00-9:05** **Règlement intérieur du Hcéres par A SEIGNEURIN**  
**Présentation du Comité par son président**  
*Séance publique (tous les membres de l'unité)*
- 9:05-9:40** **Présentation scientifique et administrative de l'unité (35 mn)**  
15 min présentation + 20 min discussion  
*Séance publique (tous les membres de l'unité)*
- 9:40-10:30** **Présentations scientifiques par les chefs d'équipe**  
15 min présentation + 10 min discussion (soit 25 min par équipe au total)  
*Séance publique (tous les membres de l'unité)*
- 9:40-10:05: Equipe Epidémiologie génétique des cancers (Andrieu) (25 min)**
- 10:05-10:30 : Equipe Biologie des systèmes du cancer (Barillot) (25 min)**
- 10:30-10:40** **Pause (10 min)**
- 10:40-12:20** **Présentations scientifiques par les chefs d'équipe**  
*Séance publique (tous les membres de l'unité)*
- 10:40-11:05 Equipe Biologie computationnelle et génomique intégrative du cancer (Cavalli) (25 min)**
- 11:05-11:30 Equipe Apprentissage statistique et modélisation des systèmes (Walter) (25 min)**
- 11:30-11:55 Equipe Pharmacologie des systèmes appliquée au cancer (Ballesta) (25 min)**
- 11:55-12:20 Equipe Méthodes statistiques pour la médecine personnalisée (Latouche) (25 min)**
- 12:20-13:30** **Déjeuner à huis clos du comité et pause (1h10 mn)**  
*Réunion à huis clos*
- 13:30-14:00** **Rencontre collective avec les représentants des établissements (30 mn)**  
*Réunion à huis clos*

- 14:00-14:25 Présentations scientifiques par les chefs d'équipe**  
*Séance publique (tous les membres de l'unité)*
- 14:00-14:25 Equipe Plateforme bio-informatique (Hupé/Servant) (25 min)**
- 14:25-14:55 Rencontres simultanées en trois groupes parallèles d'experts du comité**
- Rencontre avec les ITA (30 mn)**  
*En l'absence de tout personnel d'encadrement (directeur, chefs d'équipe)*
- Rencontre avec les chercheurs (30 mn)**  
*En l'absence de tout personnel d'encadrement (directeur, chefs d'équipe)*
- Rencontre avec les post-docs, doctorants et les étudiants (30 mn)**  
*En l'absence de tout personnel d'encadrement (directeur, chefs d'équipe)*
- 14:55-15:10 Pause (15 min)**
- 15:10-15:30 Rencontre avec le directeur et les chefs d'équipe de l'unité (20 mn)**  
*Réunion à huis clos*
- 15:30-17:00 Débriefing du comité**  
*Réunion à huis clos*

## GENERAL OBSERVATIONS OF THE SUPERVISORS

## CENTRE DE RECHERCHE

Institut Curie

### Direction

26 rue d'Ulm  
75248 - Paris Cedex 05

Alain PUISIEUX, Directeur  
Tél. : +33 (0)1 56 24 66 22  
Email : alain.puisieux@curie.fr

### U 900

Emmanuel BARILLOT  
Tél. : +33 (0)1 56 24 69 79  
Email : emmanuel.barillot@curie.fr

## Objet

Comments to HCERES Evaluation report of the Unit  
Computational Oncology – U900-CBIO  
DER PUR250024475 EV 0753172R  
Evaluation campaign 2023-2024 / Group D

### HCERES

For the attention of HCERES President,  
Mr Stéphane Le Bouler  
and the HCERES Expert Committee

Paris, 17<sup>th</sup> May 2024

Dear All,

We would like to thank the HCERES Committee for their positive comments and valuable recommendations, and all the time spent on the evaluation.

For a clarification purpose, we would like to address some issues and questions regarding the Unit, the CUBIC Platform and some teams.

### Unit:

- Page 8, part 1/, paragraph “Weaknesses” and Page 12, part 1/, paragraph “Weaknesses”: *No specific, well-defined scientific objective that would be shared by several teams beyond technical shared necessities or the access to specific skills form another team*
  - *No big common objective*
    - The common objective is understanding all aspects of tumor onset, progression, treatment and resistance, based on cutting edge numerical approaches. It is quite large and focusing on common projects often brings back to technical or focused objectives indeed (NGS, high dimension statistics, machine learning...). Nevertheless, this leaves space for a lot of interactions and collaborations as outlined in the report.
- Page 13, last line question: the patent of Walter team has been granted, the one from Ballesta team has been submitted but has not been granted so far. A second patent application from Walter team was deposited on 19/12/2022.

### Team CUBIC:

As mentioned in the evaluation report, the CUBIC team has been evaluated according to criteria adapted to research teams which explain most of the evaluation comments.

- Page 17, part “Weaknesses ...”, first two paragraphs: *“The small number of publications ... It seems that a large part of the successes of the team rest on the shoulders of one single member (...) This situation raises the question, either of an unbalanced distribution of the work within the team, or a less recognition of the work of others.”*

The activity of the CUBIC team can be broadly separated into data analysis (N. Servant), and data management/HPC (P. Hupé). While both aspects are keys for most of the projects, their recognition

is different in terms of publication (we guess that is what is meant by “successes” in the comment). People involved in data analysis are more frequently involved in the writing of the manuscripts and in the biological interpretation of the results and, as such, are usually co-authors of the publication. Data managers are sometimes acknowledged in the publication, even if we all agree that this point could be improved in the future. As the manager of the data analysis team, N. Servant is involved in most of the data analysis projects, which could explain the unbalanced contribution mentioned by the committee. Of note, some of the senior bioinformaticians of the team (E. Girard, P. Gestraud, L. Martignetti) are also regularly the only CUBIC co-authors of publications. See also the reply to Page 18 recommendations below.

- Page 17, part “Weaknesses ...”, last third paragraph: *“There is no information about the proportion of women in the team, and concrete measures to correct for a possible imbalance, as it is often the case in the field.”*

Currently, the CUBIC team is composed of 6 women and 11 men (gender ratio 0.35). Smoothed over the last 5 years, the gender ratio is 0.47 (18 women and 20 men).

- Page 17, part “Weaknesses ...”, last two paragraphs: *“Career perspectives for engineers are not mentioned and could impact the attractivity of the proposed positions. It is not mentioned if engineers are subject to a (regular) scientific evaluation of their activity, and if yes, which criteria this evaluation relies on”*

Engineers are evaluated every year as required by the Institut Curie Human Resources department. Their evaluation is based on the objectives which have been defined with their manager for the current year. The objectives can be specific to a given scientific project and/or include more transversal activities including good bioinformatics practices, training and mentoring. Career perspectives depend on their ability to manage a project, to supervise another engineer of the team, or to tackle new challenges and to move to new projects.

- Page 18, part “Recommendations...”: *“The business model and publication strategy of the platform should be clarified.”* (the reply below also applies for the above first remark in the CUBIC list)

Regarding the publication strategy of the platform, it is important to remind that its main mission is to provide support in bioinformatics to other research teams/projects. In this context, they are rarely the principal investigator of the projects they are involved in, which explains why the members of the team are usually not leading (first or last) authors. Of note, we would like to emphasize that the platform is regularly in the two first/last authors, depending on their involvement on the project. Regarding the publication of bioinformatics pipelines, from our experience, this is something which remains difficult as main of the international journals require concrete use cases, extensive benchmarkings or new biological findings highlighting the novelty of the software. Developing high-quality, reproducible, and scalable pipelines as we did in the last years, is unfortunately not enough to publish a pipeline. However, our good practices in the deployment and development of FAIR bioinformatics pipelines have been published (Kamoun et al., 2021; Allain et al. 2022).

Regarding the business model of the platform, there are four sources of funding. Today, among the 17 members of the team, 7 are paid by Institut Curie, 3 by INSERM/CNR, 4 on research grants and 3 on billing. The main source of billing today relies on the support of the sequencing facility, for which the platform team guarantees the data management, the development of pipelines, and the automatic processing of the generated data.

- Page 18, part “Recommendations...”: *“It is also unclear whether they plan for more openness to teams outside Curie (beyond public software development).”*

The platform is open to teams outside of Institut Curie following two different modalities. For standard bioinformatics questions for which they already have all tools/methods in hand, and which do not require more than 2 weeks of work, they can take the project in charge with a per day billing system. In average, this represents 2 to 5 requests per year.

On the other side, they are very open to mid/long term partnership, where they could apply to (inter)national grants as collaborators. In this case, they will ask for the funding of a bioinformaticians for a given period. Every year, they are involved in more than 10 different grant applications, for which 30% are led by external collaborators.

- Page 18, part “Recommendations...”: *The team has been evaluated according to criteria adapted to the evaluation of research entities. Some specific information is lacking, such as the number of projects submitted to the platform, decision process, project management process, user satisfaction.*

They currently do not have a well-defined platform to submit projects for the main reason that most of their mid/long-term projects are funded through research grants. As soon as they have the technical expertise to support a project, or if they think that being part of a project could be a strategic added value for the team, they always try to participate. Then, the selection is performed depending on the funding (or not) of the project. Regarding short term/standard analysis projects, with only a few requests per year, they currently do not need a submission platform. The low number of standard projects can be explained by the fact that most of the research teams at Institut Curie now have their own bioinformaticians, with a strong link to the platform through the bioinformatics hub.

Regarding the user satisfaction, the platform has a steering user committee which meets every year. This committee is made of principal investigators and managers from the Research Center and the Hospital with strong needs in bioinformatics. They participate in ensuring a connection between all potential users and the core facility (thus reporting any (un-)satisfaction), fixing the main orientations of the platform for the next year and write some recommendations which are then presented to the direction of the institute. When needed, they organise surveys on specific topics to collect needs or evaluate satisfaction (the last one being on the functioning of the bioinformatics hub in 2023).

#### **Team Lesueur :**

Team 1 will maintain and will even try to increase its role in breast cancer risk research and pursue new research fields such as on thyroid or pediatric cancers. To ensure the achievement of the projects and the maintenance of the national studies and PIGE effort, the ongoing effort to answer calls including EU calls, will be pursued.

- *Recommendation 3: “The number of publications of PhD students as first author is low, two PhD students among the three who have defended their thesis, having a single paper as first author.”*

The team recognizes that the first-author publication rate of students who have defended their thesis is low. According to recommendations of doctoral schools, the thesis proposed projects are initially designed so that they can lead to the publication of three articles as first author. These subjects have sometimes evolved into the prioritization of an ambitious research question involving the use of data generated in national or international consortia. (one consequence of such work is the extended publication time given the numerous partners involved (e.g. Juliette Coignard’s Nature Communication paper). In general, first author publications take time but are obtained finally with all the works during the thesis.

**Team Ballesta:**

Page 30 *"the team depends on external collaborations to obtain data"*.

Since its creation in 2018, the team generates their own experimental data, both in vitro and in vivo in mice, which yielded several recent publications (e.g. Dulong et al. 2022 PMID: 36432655 and Catozzi et al. PMID: 36537068). One of the permanent researchers, S. Dulong, is a highly trained chronobiologist while the team PI also performs in vitro experiments herself. We currently use experimental facilities from our former location (Hôpital Paul Brousse, Villejuif) and will move our wet lab activity to Saint-Cloud when the laboratories w available. We also rely on technological platforms of Institut Curie (ongoing proteomics and single-cell study).

We thank you once again for your attention on our report and hope we could clarify some aspects for the understanding of all.

Yours Sincerely,

Pr Alain PUISIEUX  
Directeur du Centre de Recherche de l'Institut Curie



Dr. Emmanuel BARILLOT  
Directeur de l'U 900



The Hcéres' evaluation reports are available online:  
[www.hceres.fr](http://www.hceres.fr)

**Evaluation of Universities and Schools**  
**Evaluation of research units**  
**Evaluation of the academic formations**  
**Evaluation of the national research organisms**  
**Evaluation and International accreditation**



2 rue Albert Einstein  
75013 Paris, France  
T.33 (0)1 55 55 60 10

[hceres.fr](http://hceres.fr)

 [@Hceres\\_](https://twitter.com/Hceres_)

 [Hcéres](https://www.youtube.com/Hceres)

