Processus de sélection - AAP générique 2018 étape 2

MUPIDE-4D

Coordinateur du projet

Nom : **TATOUT**
Prénom : **Christophe**

Aucun droit de réponse n'a été enregistré par le coordinateur de projet.

1 - EVALUATION : 65833

QUALITE ET AMBITION SCIENTIFIQUE

- CLARTÉ DES OBJECTIFS ET DES HYPOTHÈSES DE RECHERCHE
- CARACTÈRE NOVATEUR, ORIGINALITÉ, PROGRÈS PAR RAPPORT À L'ÉTAT DE L'ART
- FAISABILITÉ NOTAMMENT AU REGARD DES MÉTHODES, GESTION DES RISQUES SCIENTIFIQUES

The applicants seek to develop an integrated bioinformatics platform for the study of the higher order organization of the genome. The research team proposes to first develop several computational platforms to analyze, in a coordinated fashion, imaging data and molecular mapping data or omics data. They will demonstrate proof of principle of the usability of the platform and apply it to the study of neuronal progenitor differentiation and will, finally, using the platform generated in aim 1 and the data from aim 2, derive a comprehensive combined imaging-omics based model of the process.

This proposal focuses on the problem of nuclear organization and genome architecture which is a rapidly expanding important field. The question of how genomes are organized and how they contribute to function is an important one, that has recently been put on the map and is currently being very actively explored. The proposed studies are a potentially important contribution to this field.

The field is based on two types of approaches, imaging and omics (RNAseq, Hi-C etc.), however, these diverse datasets are generally analyzed separately. The proposed approach fills the important niche of attempting to combine the two types of data, which ultimately will be needed to make progress. The proposed efforts are important.

The proposed approaches are sound. Many aspects of the strategy are based on already available tools or their modification which increases confidence of success. The key to success is the proposed development of compatible computing pipelines for both imaging and omics using a common programming language and integrated platforms. While challenging, this is not an overly high-risk project, it simply requires the dedicated and concerted development of the proposed tools.

The major contribution from this work will come from aim 1 which represents the actual development of the computational platform. Aim2 will generate the necessary data to implement and test the timeline. Aim 3 will demonstrate the usefulness of the approach.

A weakness of the proposal is the exclusive use of data generated in aim 2. Considering that the main goal of the project is to generate a widely usable computational platform, it would seem essential to test the pipeline not exclusively on in-network generated data of a particular type, but rather it would be important to test the pipeline on publicly available data or data from various sources. If the pipeline can only be applied to in-house generated data, its value will be diminished. Demonstration of its applicability to imported, including publicly avaialable, datasets would have been essential.

Another weakness is the absence of analysis of Hi-C data. Rather the applicants only use 4C data and the somewhat obscure High-Salt Recovered Sequence assay. Hi-C is now fairly standard and should be included in this analysis. The types of data analyzed here are somewhat outdated.

Another weakness is the exclusive use of gene position as an imaging read out. This will not tell us very much since its relevance to 4C will not be clear. It would have been important to probe by imaging pairs of interactors identified by 4C, or even better by Hi-C.

ORGANISATION DU PROJET ET MOYENS MIS EN ŒUVRE

- COMPÉTENCE, EXPERTISE ET IMPLICATION DU COORDINATEUR SCIENTIFIQUE ET DES PARTENAIRES **- QUALITÉ ET COMPLÉMENTARITÉ DU CONSORTIUM, QUALITÉ DE LA COLLABORATION** **- ADÉQUATION AUX OBJECTIFS DES MOYENS MIS EN ŒUVRE ET DEMANDÉS**	The investigators appear well positioned to perform the proposed studies. The team is diverse in that it is composed of computational experts and biologists. All necessary expertise appears to be presented in the consortium. This is a group of productive, well established investigators. All investigators are well qualified to contribute to the team. The team leader has experience in managing larger groups and a clear strategy for the coordination of the group has been described. The tasks for the various investigators are well defined and well-coordinated. The requested resources are appropriate for the proposed work.

IMPACT ET RETOMBEES DU PROJET

- CAPACITÉ DU PROJET À RÉPONDRE AUX ENJEUX DE RECHERCHE DÉFINIS DANS LE PLAN D'ACTION **- IMPACT POTENTIEL DANS LES DOMAINES SCIENTIFIQUE, ÉCONOMIQUE, SOCIAL OU CULTUREL** **- STRATÉGIE DE DIFFUSION ET DE VALORISATION DES RÉSULTATS**	If successful, this work will have a major impact. One of the major needs in the field of higher order genome organization is the development of an integrated computational pipeline. Similar efforts are under discussion as part of the NIH-funded 4D Nucleome project currently taking place in the United States ($120M over 5 years, started in 2016, possible extension for another 5 years), however, these are currently not successful due to the very large and diverse group of investigators in the consoritum. The focused effort proposed here involving a smaller, closely collaborating group of investigators appears a much more promising approach. If successful, this pipeline may become the standard for combined imaging-omics multi-parameter analyses in the entire community. The impact of the biological insight gained from these proposed studies is considerably lower than the methodology development aspect of this proposal. Because sub-optimal parameters (gene position, 4C) will be analyzed rather than more functionally relevant parameters (interaction by FISH, hi-C) only limited biological information will be generated. The impact of the work is limited by a relatively minimal plan for dissemination of the results. The only dissemination will be through publications, which of course is appropriate, and summers schools and local communication to students in the involved laboratories. If successful it would be important for example to communicate these to the 4D Nucleome consortium and to highlight them at international scientific meetings.

AVIS GENERAL

This is an excellent proposal which addressed an important biological problem and fills an urgent need in an expanding and exciting field of research. The value and major potential contribution of the proposal is the generation and dissemination of an integrated computational platform for the analysis of imaging and omics data. The proposal is weakened by its focus on somewhat outdated methods (4C, HRS) and the exclusive use of in-house generated data rather than diverse sets of datatypes; this will limit its potential use in the community.

2 - EVALUATION : 70606

QUALITE ET AMBITION SCIENTIFIQUE

- CLARTÉ DES OBJECTIFS ET DES HYPOTHÈSES DE RECHERCHE
- CARACTÈRE NOVATEUR, ORIGINALITÉ, PROGRÈS PAR RAPPORT À L'ÉTAT DE L'ART
- FAISABILITÉ NOTAMMENT AU REGARD DES MÉTHODES, GESTION DES RISQUES SCIENTIFIQUES

The grant “MUPIDE-4D Multi-Platform Interoperability Dedicated to the 4D-nucleome” suggests to develop a multiscale component platform to integrate data coming from genomics, transcriptomics, proteomics with microscopy images to enable bioinformaticians, biostatisticians, and biologists to correlate data coming from these different technologies. This main goal is to help in the launching of the European 4D nucleome project.

While a unifying tool for data integration is much needed, I think the approach suggested by the researchers is somewhat lacking. The main weakness of the proposal is that it lacks a physical model to integrate all the accumulated data. While the authors do suggest to use polymer model to account for some of the data (4C experiments), they do not seem to treat this physical representation of the chromosomes as the model onto which the omics and transcription data will be projected. Thus, we are left with correlations and statistical analysis as the main tool of research. I am not sure that the suggested project would fully use the data coming from imaging approaches.

The bulk of the project lays in the integration of many different experimental approaches to study transcription and differentiation of mouse embryonic cells into Neural Progenitors. Since most of the experimental techniques have already been developed, the main risk element here seems to be (according to the authors) the use of Octrees as the main structure to represent the resulting high dimension data. A priory, it is not obvious why this specific object is the best approach to close the gap between chromosome capture results, 3d positioning of chromosome loci, transcription levels, histone modification data, and differentiation markers. Perhaps this point could be justified.

ORGANISATION DU PROJET ET MOYENS MIS EN ŒUVRE

- COMPÉTENCE, EXPERTISE ET IMPLICATION DU COORDINATEUR SCIENTIFIQUE ET DES PARTENAIRES - QUALITÉ ET COMPLÉMENTARITÉ DU CONSORTIUM, QUALITÉ DE LA COLLABORATION - ADÉQUATION AUX OBJECTIFS DES MOYENS MIS EN ŒUVRE ET DEMANDÉS	• The new aspect which the authors suggest to integrate to enable discovery is data coming from imaging. They plan to concentrate on the positioning of 10 loci on the chromosome of mouse embryonic stem cells. The authors suggest imaging these loci at three time points (day0, day 8 and day 12) using DNA FISH. The authors further suggest studying local chromatin compaction next to these loci. The authors plan to perform chromosome capture (CC) experiments centered on these loci. What is lacking, is how these data will be integrated. It has already been demonstrated that the 3d positioning of genomic loci correlates with the results of CC experiments. Thus, results from these different approaches would corroborate one another. However, the novelty here should come from a new approach to integrate these data within a single model to learn something new, or to reduce the error rate resulting from the use of every single experimental technique. • The authors discuss the contribution of the platform to study the dynamics of genome organization. However, the actual dynamics of chromosomes is almost not addressed in the proposal. Indeed, the dynamics of chromatin has been shown to be important during cell differentiation [Masui et al., Cell 145, 447 (2011)]. In recent years it is increasingly easier to use Cas-9 the study of chromatin loci motion by fluorescent imaging [Maass et al., Nature Structural & Molecular Biology 25, 176 (2018)]. Since the authors suggest to develop the platform over the next four years, it would be very important not to disregard this essential aspect of chromatin functioning. This aspect of genome function can only be studied using imaging techniques. Hence, it should be central to the proposal. Here, the use of physical models of polymer and stochastic analysis will be important to extract new features and study them in combination with the other components (omics, CC, etc.). Currently, CC experiments are only snapshots of the 3d organization of chromatin. Thus, making use of imaging approaches could bridge this gap between the CC and microscopy. • The authors plan to begin by designing the database structure for metadata collection, and the development of R packages. It may be beneficial to begin by clearly defining the novel parameters which are unique for imaging. For example, when treating movies, will the platform make use of a tool such as ICY to perform tracking? Which statistics of the nuclear geometry are most important? In this respect perhaps tasks 5 and task 2 are too distant in their scheduling.

IMPACT ET RETOMBEES DU PROJET

- CAPACITÉ DU PROJET À RÉPONDRE AUX ENJEUX DE RECHERCHE DÉFINIS DANS LE PLAN D'ACTION - IMPACT POTENTIEL DANS LES DOMAINES SCIENTIFIQUE, ÉCONOMIQUE, SOCIAL OU CULTUREL - STRATÉGIE DE DIFFUSION ET DE VALORISATION DES RÉSULTATS	• The authors suggest deploying their package as an ImageJ plugin or R packages. While the bioinformaticians and biostatisticians communities often use R, its use is less frequent amongst biophysicists and biologists. In such a multi-scale platform, it may be beneficial to be able to use Matlab or python to write scripts for data analysis. • Given that at least one aspect of the platform would result in a polymer model (Tasks 7,8), it would beneficial if this model could then be used for simulations with a tool such as Lammps. This approach would allow asking more biophysical questions. For example, what would be the physical -dynamics model that would a chromosome configuration at day 0 to evolve to another configuration at days 8 and 12.

AVIS GENERAL

To conclude, I would suggest that the authors would take a more biophysical approach to integrating data from the different experimental techniques. Statistical analysis and multivariate analysis has proven very successful in describing cell identity. However, to make full use of the type of data originating from imaging experiments, it would be beneficial, in my view, to take a more physical-model specific approach.