NOVADISCOVERY |INTRODUCTION

You are an engineering student curious about sciences, particularly in the field of systems biology. You are open-minded and ready to work with a team of innovation enthusiasts in various fields of expertise. At Novadiscovery, you will participate in the development of *in silico* biomedical models with direct applications to improve new drug R&D.

Novadiscovery is a pioneer in the emerging field of *in silico* medicine based in Lyon (France). Working with an innovative company in its scale-up phase is an entirely different experience than working for a large established company. You will be assigned a great number of responsibilities and work in a dynamic environment with strongly motivated people who will help you fast-climb a steep learning curve. For more information visit www.novadiscovery.com

YOU ARE…

- ➔ **A team player,** a good listener, and an effective communicator
  Join a growing multidisciplinary team of enthusiastic innovators
- ➔ **Curious and proactive** with a solid grounding in biology
  Particularly in cell biology, molecular biology, and omics, to address real-life clinical issues.
- ➔ **Autonomous** and self-motivated with strong analytical and problem-solving skills
  Find innovative solutions to science and engineering problems
- ➔ **Eager to learn** and use mathematical methods for the modeling of biological systems
  Simulate virtual diseases and treatments with ODE, PDE, Monte-Carlo Simulations
- ➔ **Willing** to explore and exploit large datasets and virtual populations
  Apply machine learning, statistical analysis, and outliers detection
- ➔ **Responsive** and capable of facing time-sensitive challenges
  Project management with client-facing opportunities are awaiting you

YOU WILL…

- ➔ **Contribute** actively to the creation of *in silico* pathophysiological models
- ➔ **Impact** the development of the company's simulation platform
- ➔ **Analyze** and exploit large simulation results
- ➔ **Participate** in weekly and monthly project meetings and reporting

TECHNOLOGIES & LANGUAGES

We are looking for people who know some of the following or are eager to learn and work with them:

- Unix environment (MacOSX, Linux)
- Programming languages (Haskell, Python, NodeJS)
- Statistical/scientific computing (R, Matlab)
- Big data (SQL, Spark)
- Markup languages (Markdown, LaTeX)
- Miscellaneous (Git, bash, zsh, RDF, SPARQL, Elasticsearch)

DETAILS

Type	Internship
Salary	Competitive
Start date	**Flexible**
Contact	recruitment@novadiscovery.com

APPLY

Online Form

SUBJECT MATH/DEV 1 | SBML Solver Benchmarking

Background:

The Systems Biology Markup Language (SBML) is the de facto standard to encode computational models of biological processes in a machine-readable format. Multiple softwares have been written to read/write from such format and to run simulations on those models. Novadiscovery has implemented its own version of SBML parser/writer and model solver.

Objective:

Develop a comprehensive benchmarking tool to compare:

- The richness of the features supported by each software
- The speed of the solver
- The validity of the results generated

Work Process:

- Investigate which softwares are programmatically accessible
- Implement a reproducible environment (via a virtual machine, docker or nix shell) where all those softwares are available
- Implement a benchmarking framework (in Python or Haskell) that loads a predefined set of models from the Biomodels Database, solves them and compares timing and results

Deliverables: SBML benchmarking framework

Technologies & languages: SBML, Nix, Docker, Haskell, Python

Keywords: Biomodels, Systems Biology, Performance analysis, Reproducible environment

SUBJECT MATH/DEV 2 | Model Stability Analysis & Active Sub-Spacing of the Input Parameters

Background:

Systems biology models usually have a high degree of complexity with many free parameters. Measuring those parameters in vivo is sometimes difficult or even impossible, so we are left to calibrate them to observable data. However, the parameter identification process can be problematic when models are "sloppy" - see [2,3]. The goal of this internship is to implement an algorithm to automatically detect whether a model is sloppy and try to compress the parameter space dimension via active sub-spacing - see [1].

Objective:

Detect sloppiness automatically and implement methods for parameter space reduction and/or model constraint automation

Work Process:

- Conduct a literature review
- Implementation of a parameter sensitivity computation module [need to be a parallelizable code given the high dimensionality of the problem]
- Implement a parameter space reduction algorithm

Deliverables: Dedicated Modules and functionalities in our simulation platform

Technologies & languages: Haskell, Python

References:

[1] https://arxiv.org/pdf/1303.6738.pdf
[2] http://www.lassp.cornell.edu/sethna/Sloppy/SloppyPubs.html
[3] http://www.lassp.cornell.edu/sethna/pubPDF/SloppyEverywhere.pdf

Keywords: Applied Mathematics, Sensitivity Analysis, Systems Biology, Inverse Problems

SUBJECT MATH/DEV 3 | Global Optimization Algorithm for Virtual Population Design

Background:

Novadiscovery evaluates the outcomes of treatments by applying them *in silico* to a population of virtual patients. Those virtual patients are represented by a vector of descriptors that impact the model behavior. Thus, the same model can behave differently from one patient to the next, similar to real life situations when the outcome of a treatment varies from one individual to the next. Unfortunately, for a lot of those descriptors, their exact value or value distributions are unknown from the onset. Being able to infer those from indirect information such as expected behaviors of our pathophysiological model or observable data points from in vivo or in vitro experiments then becomes a crucial exercise so as to simulate our models on realistic patients.

Objective:

- Implement and compare several optimization algorithms to try and calibrate unknown descriptor distributions.
- Explore genetic algorithms, as they lend themselves naturally to the problem of finding a population of realistic patients

Work Process:

- Literature review
- Implementation of a simple evolutionary optimization algorithm
- Calibration of a virtual population with desired properties

Deliverables:

- Target functions for the evaluation of fitness of individuals
- Implementation of evolutionary operations
- An operational algorithm to minimize target function

Technologies & languages: Distributed Computing (AWS), Docker, Python, Celery, Haskell

References:

[1] https://en.wikipedia.org/wiki/Evolutionary_algorithm
[2] http://www.sciencedirect.com/science/article/pii/S0377042713006274
[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4809626/pdf/PSP4-5-140.pdf

Keywords: Evolutionary Algorithm, Dynamic Patients, Systems Biology

SUBJECT MATH/DEV 4 | NLP - Relation Extraction from Life-Science Literature Applied to Logical Modeling

Background:

GitHealth is an online platform developed by Novadiscovery to assist biomodelers in literature review and knowledge model edition. To enhance knowledge review in GitHealth, Natural Language Processing (NLP) techniques are being explored, including Named Entity Recognition (NER) and Relation Extraction (RE).

Objective:

- Implement a service in GitHealth to identify relations between biological entities in scientific articles: eg. **Protein A** *interacts with* **Protein B**, as part of the structured knowledge extraction process
- Investigate the resulting relation tree through graph reasoning, combined with relations extracted from Semantic Web knowledge bases

Work Process:

Based on the existing work on NER at Novadiscovery:

- Conduct a literature review on RE. Identify the technical uncertainties, pitfalls and specificities of the use of NLP applied to life-science literature
- Implement a Relation Extraction engine
- Construct a logical model based on fuzzy logic to explore completeness of the knowledge model
- Integrate the NLP engine as a web service in GitHealth.

Deliverables:

- Relation Extraction engine
- Logical model generator

Technologies & languages: Python, RDF, Sparql, NodeJS, Elasticsearch

References:

[1] An Unsupervised Text Mining Method for Relation Extraction from Biomedical Literature (link)
[2] GENIA corpus—a semantically annotated corpus for bio-text mining (link)
[3] Enhancing Boolean networks with fuzzy operators and edge tuning (link)
[4] Logml Logical Modeling Tool (link)

Keywords: NLP, Named Entity Recognition, Relation Extraction, Semantic Web, GitHealth

SUBJECT MATH/DEV 5 | GitHealth - Enhancement of the Knowledge Model Edition Tool

Background:

GitHealth is an online platform developed by Novadiscovery to assist biomodelers in literature review and knowledge model edition. Currently missing features include the simultaneous (by several biomodelers at a time) editing of a document, as well as the possibility to comment and suggest modifications on shared documents (Google Drive like).

Objective:

Contribute to the development and integration of new modules into a web-based application. (The internship may also include working on other projects for client-side or server-side use.)

Work Process:

- Familiarize with the role of knowledge modeling in Novadiscovery workflow for *in silico* clinical trial modeling and GitHealth's user needs
- Review existing tools, techniques and design patterns
- Implement a discussion, comment and notification module (Javascript & Node.js)
- Implement a module to allow the simultaneous edition of text documents (for instance knowledge models) in GitHealth

Deliverables: Dedicated Modules and functionalities for GitHealth

Technologies & languages: NodeJS, VueJS / Vuex, Websocket

References:

[1] Etherpad editor (The foundation for Google Doc editor) (link)
[2] Lessons learned from creating a rich-text editor with real-time collaboration (link)
[3] Automerge for collaborative text editing (link)

Keywords: Front-end, Full Stack, GitHealth, WebSocket, Real-Time Collaborative Editor

SUBJECT BioMod 1 | SysCLAD project

Background:

SysCLAD was a two-year project funded by the European Union in 2014 which aimed towards developing a mathematical model to predict whether recipients were at risk of developing chronic lung allograft dysfunction (CLAD) in the 3 years succeeding a lung transplant. For this internship project, you will implement the initial model and develop two additional biological submodels (lung fibrosis & immune system) to compare the efficacy of two possible treatments. The results will be used to prepare at least 1 publication.

Objective:

Develop two computational submodels (lung fibrosis and immune system) on the basis of the work already done (complete literature review concerning these two biological mechanisms)

Work Process:

- Update (if needed) the literature review of the two biological models via Novadiscovery's knowledge modeling framework (GitHealth)
- Create two computational models via Novadiscovery's simulation framework (Simwork)
- Integrate the two new submodels with the previous ones
- Redact an *in silico* experimental protocol and run simulations to compare the effect of the tested treatments
- Redact a scientific article (to be submitted)

Deliverables:

- Two calibrated computational models
- *in silico* experimental protocol
- Simulations report
- Scientific article

Technologies & languages: Internal Tools (GitHealth and Simwork) for creating the models, R for result analysis.

References:

[1] Royer PJ et al. Chronic Lung Allograft Dysfunction: A Systematic Review of Mechanisms. Transplantation: September 2016 - Volume 100 - Issue 9 - p 1803–1814. Doi: 10.1097/TP.0000000000001215
[2] DerHovanessian A et al. Chronic Lung Allograft Dysfunction: Evolving Concepts and Therapies. Semin Respir Crit Care Med. 2018 Apr;39(2):155-171. doi: 10.1055/s-0037-1618567

Keywords: Biomodeling, Lung Allograft, Systems Biology

SUBJECT BioMod 2 | New drug R&D

Background:

Novadiscovery is a pioneer in the field of *in silico* clinical trials, which are poised to become an industry standard as regulators now see Modeling and Simulations (M&S) as a strategic priority. Each commercial project is aimed to unlock the potential of M&S and allow our biotech and pharma partners to accelerate and de-risk the R&D of new therapies by establishing their clinical benefits upstream of human trials. To predict drug efficacy, Novadiscovery applies a proprietary methodology (the Effect Model) with WISE® (Whitebox In Silico Engine), an open ecosystem which brings together the modeling and simulation expertise of the company.

Objective:

- Contribute to the creation of pathophysiological models and the exploitation of their results.
- Develop a submodel which will be integrated into a complete model. The actual model to be implemented will depend on the partner's need of the moment (The intern will be part of the R&D process within that project).

Work Process:

- Participate in weekly and monthly project meetings and reporting (scientific and project management meetings)
- Conduct literature reviews on the biological system to model
- Apply logical modeling of the physiopathological system
- Create a computational model in the simulation platform
- Integrate the new submodel in the complete model
- Redact an *in silico* experimental protocol and run simulations to answer the client's questions

Deliverables:

- Knowledge Model
- Logical Model
- Computational Model

Technologies & languages: Internal Tools (GitHealth and Simwork) for creating the models, R for result analysis.

References: To be provided based on the drug candidate chosen

Keywords: Commercial Project, Systems Biology, Biomodelling, Drug R&D