# Content:

Describe how the scientific environment in which the research will be done contributes to the probability of success (e.g., institutional support, physical resources, and intellectual rapport). In describing the scientific environment in which the work will be done, discuss ways in which the proposed studies will benefit from unique features of the scientific environment or from unique subject populations or how studies will employ useful collaborative arrangements.

If there are multiple performance sites, describe the resources available at each site.

Describe any special facilities used for working with biohazards and any other potentially dangerous substances. Note: Information about select agents must be described in the Research Plan, Select Agent Research.

For early stage investigators (ESIs), describe institutional investment in the success of the investigator. See NIH's New and Early Stage Investigator Policies. Your description may include the following elements:

• resources for classes, travel, or training

• collegial support, such as career enrichment programs, assistance and guidance in the supervision of trainees involved with the ESI’s logistical support, such as administrative management and oversight and best practices training;

• logistical support, such as administrative management and oversight and best practices training;

• financial support, such as protected time for research with salary support.

Include a detailed description of the institutional facilities and resources available to the fellowship applicant. The information provided is of major importance in establishing the feasibility of the goals of the fellowship training plan.

# FACILITIES & OTHER RESOURCES

Institution

The CSHL main campus is located on 116 acres on the north shore of Long Island, approximately 35 miles east of Manhattan.

Dr. Desmarais is in the Kinney lab which is part of the Simons Center for Quantitative Biology (SCQB), a vibrant program located in the Hillside Building Complex on the CSHL main campus. The SCQB runs weekly seminars and journal clubs that focus on topics in genomics, with an emphasis on quantitative sequence-function relationships and deep learning. These seminars and journal clubs provide both a training opportunity to expand and deepen knowledge of these topics and a venue for building collaborations with other scientists in SCQB. Dr. Desmarais regularly attends these events and will continue to do so through the course of this project. In addition to these events, The Kinney lab meetings are held jointly with the Koo and McCandlish labs. The Koo lab specializes in using interpretable deep learning approaches in genomics while the McCandlish lab specializes in applying theory to understanding evolution. By attending these meetings Dr. Desmarais will continue to build his knowledge of these subjects and his network of collaborators.

The CSHL Meetings and Courses Program provides further opportunities for collaboration and exposure to cutting edge research, factors that will contribute to the success of the proposed studies. In particular, scientists at CSHL have numerous opportunities to exchange ideas and develop collaborations with other leading researchers through CSHL’s unique and world-renowned forum for meetings and courses, with over 9,000 scientists participating each year.

Laboratories

The Kinney laboratory is located on the CSHL main campus in the Hillside Building Complex. The Kinney Lab includes a **dry lab** consisting of one 95 sq. ft. office with room for 2 trainees, 437 sq. ft. of shared office space for additional trainees, and a 274 sq. ft. **wet lab*,*** located 1 floor below the offices, with bench and desk space for 4 experimentalists. There is a nearby common area that includes space for group meetings and discussions, audio-visual equipment for presentations and remote conferencing, network printers, and a photocopier.

Shared Resources

CSHL provides several state-of-the art shared resource facilities to support the research of their faculty. Available facilities pertinent to the proposed project include:

* NextGen DNA Sequencing Core Facility
* Sanger DNA Sequencing Core Facility
* Flow Cytometry Core Facility
* Antibody Core Facility
* Tissue Culture Facility
* Mass Spectroscopy Core Facility
* Machine Shop
* Bioinformatics Core Facility
* High Performance Computer Cluster

Computational Resources

Dr. Desmarais has access to an institutionally shared High Performance Compute Cluster (HPCC). This 2528 core compute cluster includes a BlueArc Titan 3210 high-performance storage system housed in the 3000 sq. ft. state-of-the-art datacenter at CSHL for large-scale software testing and research. The cluster has 28 standard nodes (each with 48 cores across 2 2.1 GHz Intel Xeon 6252 processors and 769 GB of RAM) and 4 high memory nodes (each with 3 TB of RAM and 96 cores across 4 2.4 GHz Intel Xeon 8260 processors). This system supports Sun Grid Engine and Hadoop execution, including a persistent HDFS across the nodes (400TB). The machines are connected via gigabit Ethernet to the central data storage environment, with total storage capacity for >4 Petabytes (PB) that is expected to grow significantly over the coming years.

The HPCC has recently been upgraded to include four Graphics Processing Unit (GPU) nodes, one large-memory node, and one top-of-rack 10GbE switch to interconnect the servers with each other and with the rest of the high-performance computer cluster (HPCC). The fourteen GPU nodes each contain two 20-core 2.5 GHz Intel Xeon-Gold 6248 Cascade Lake processors, four Nvidia Tesla V100, SXM2, 32GB GPUs, and 768 GB RAM (DDR4-2933 MT/s). Each SXM2 GPU module supports up to six second-generation Nvidia NVLink connections for a total bandwidth of 300 GB/s, 10x faster than PCIe Gen 3, greatly accelerating inter-GPU communication. The large-memory node contains four 20-core 2.5GHz Intel Xeon-Gold 6248 Cascade Lake processors and 3 TB RAM (DDR4-2933 MT/s). Together, these nodes contribute an additional 240 CPU cores, 32 GPUs, and 6 TB RAM to the previous HPCC. For compatibility with the rest of the HPCC and most computational biology software, these nodes employ the Linux operating system. The 10GbE (Gigabit/second Ethernet) switch ensures sufficient bandwidth to handle inter-node and file-system traffic for all cores on all nodes without bottlenecks.

The Kinney lab receives support from the Information Technology Department at CSHL. This service includes network connectivity at up to 1 Gbit/sec, email accounts, tape backups and climate-controlled rack space for the cluster and the data servers. The IT Group also provides site licenses for a wide range of software and project management tools.

Other

The Kinney Lab receives administrative support from CSHL, including from two on-site and **extremely helpful** administrative assistants**.**