|  |  |
| --- | --- |
| CNMC_485_k.eps |  |
| 111 Michigan Avenue, N.W.  Washington, DC 20010-2970  202 476-6011 |  |

**DATE**

**NAME**

**DEPARTMENT**

**ADDRESS**

**RE:** Letter of Support for Proposal to the **XXXXX** titled **“XXXXXX”**

Dear Dr. X,

This letter is to formally acknowledge that the Children’s Bioinformatics Unit at the Children’s National Medical Center is pleased to provide all the bioinformatics support for the XXXXXX  proposed in your grant application cited above.

Our unit provides investigators with core bioinformatics and computational biology expertise to manage and analyze large molecular datasets like RNA, Whole Genome and Exome, T-Cell Receptor, Methylation, Chip sequencing etc, apart from regular array type analysis. Our aim is to assist investigators in the development of customized pipelines to provide solutions to hypothesis-driven biological experiments and derive meaningful results from exploration of large data sets. Next-generation sequencing data is the focus of most of our projects, but we do also encounter other biological datasets all the time. Our emphasis is on genomics and systems biology with an overarching goal of personalized medicine. We also provide expertise in statistical analysis and training where applicable.

The CBU was established in November 2017, led by Dr. Eric Vialin, Director, Center for Genetic Medicine Research and Dr. Hiroki Morizono, Director, CBU with generous support from Dr. Vittorio Gallo, Chief Research Officer, Center for Translational Science, Clinical and Translational Institute at Children’s National (CTSI-CN) and the District of Columbia Intellectual and Developmental Disabilities Research Center (DC-IDDRC) at [Children’s National Medical Center](https://childrensnational.org/). Our staff works closely with the Genomics Core at Children’s and other bioinformatics-oriented research groups at George Washington University in the development of new algorithms and software tools.

We look forward to working with you and are confident that we will provide the highest quality data analysis to help answer important research questions. Please feel free to email us (bioinformatics@chilrensnational.org), if you have any questions.

Sincerely,

Dr. Hiroki Morizono

Director

Children’s Bioinformatics Unit

**Website -** [**https://childrensnational.org/research-and-education/research-resources/informatics/bioinformatics-unit**](https://childrensnational.org/research-and-education/research-resources/informatics/bioinformatics-unit)

The CRI Bioinformatics Unit (CBU) provides bioinformatics consultation and analytical services in an initiative sponsored by the [Center for Genetic Medicine Research](https://childrensnational.org/research-and-education/center-for-genetic-medicine-research), the [Center for Translational Research](https://childrensnational.org/research-and-education/center-for-translational-research), the Clinical and Translational Institute at Children’s National ([CTSI-CN](https://www.ctsicn.org/)), the District of Columbia Intellectual and Developmental Disabilities Research Center ([DC-IDDRC](https://childrensnational.org/research-and-education/center-for-neuroscience-research/dc-iddrc)) and the [Chief Research Officer’s office](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/vittorio-gallo) at Children’s National.

We assist in the following experimental design consultation, grant technical writing support and data analysis.

Data analysis includes:

* RNA sequencing
  + Differential gene expression
  + Single nucleotide variation detection
  + Fusion analysis
* Whole genome sequencing
  + De novo assembly
  + Single nucleotide variation detection
  + Copy number variation
  + Structural variant detection
* Exome sequencing
  + Single nucleotide variation detection
  + Copy number variation
  + Structural variant detection
* Microbiome analysis (16s and metagenomics)
* Single cell sequencing
* T-cell receptor sequencing
* Methylation sequencing (bisulphite sequencing)
* Chip sequencing
* Microarray
* Other/custom analysis

The CBU also actively participates in training the next generation of biomedical researchers with the goal of enabling quality research and education through seminars and workshops.

To initiate a project with the Bioinformatics Unit, please [fill out the submission form](https://cri-datacap.org/surveys/?s=3EJP7L8PLK) to share details of your experiment and schedule a consultation.

Preliminary data processing and analysis turnaround time is typically 2-3 weeks depending upon experiment and number of samples. Complex downstream data analysis may take more time. Completion timelines will be discussed during the consultation.

# **Meet the Team**

## Leadership

* [Eric Vilain, M.D., Ph.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/vilian-eric) (Director of Center for Genetic Medicine Research)
* [Lisa M. Guay-Woodford, M.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/lisa-guaywoodford) (Director of CTSI at Children’s National)
* [Hiroki Morizono, Ph.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/hiroki-morizono) (Director of CBU)
* [Kazue Hashimoto-Torii, Ph.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/kazue-hashimototorii) (Center for Neuroscience Research liaison)
* [Susan Koblach, Ph.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/knoblach-susan) (Director of Genomics Core)
* [Michael Keller, M.D.](https://childrensnational.org/visit/find-a-provider/michael-keller) (Center for Cancer and Immunology liaison)
* [Marius George Linguraru, D.Phil., M.A., M.Sc.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/marius-george-linguraru) (Sheik Zayed Institute for Surgical Innovation liaison)

## Faculty

* [Hayk Barseghyan, Ph.D.](https://childrensnational.org/research-and-education/about-cri/staff/hayk-barseghyan)
* [Seth Berger, M.D., Ph.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/berger-seth)
* [Wei Li, Ph.D.](https://childrensnational.org/research-and-education/about-cri/faculty-and-leadership-directory/li-wei)

## Staff

* [Payal Banerjee](https://childrensnational.org/research-and-education/about-cri/staff/payal-banerjee), M.S
* [Surajit Bhattacharya, Ph.D.](https://childrensnational.org/research-and-education/about-cri/staff/surajit-bhattacharya)

**Authorship**

Per the guidelines of the International Committee of Medical Journal Editors (ICMJE)[1], authorship credit is indicated when the following four conditions are met:

* substantial contributions to: the conception or design of the work; or the acquisition, analysis or interpretation of data for the work
* drafting the work or revising it critically for important intellectual content
* final approval of the version to be published
* agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved

Therefore, members of the CRI Bioinformatics Unit who contribute substantially to study design, implementation, analysis and/or interpretation should be offered the opportunity to earn authorship on all abstracts, presentations and manuscripts stemming from joint collaborations with investigators seeking support. Investigators should invite their collaborators in the CRI Bioinformatics Unit to participate in drafting and revising abstracts and manuscripts.  
  
Authorship order should be discussed prior to the initiation of the collaboration. The order tends to vary and often depends upon the specific contributions of the collaborating bioinformatician(s). For example, it is customary for faculty bioinformaticians who have participated in the design, analysis, conduct and manuscript preparation of a clinical trial to be listed as second author. For studies requiring an intensive or high-end bioinformatics analysis, it is also customary to list the faculty bioinformatician(s) among the first three authors. Staff bioinformaticians are typically listed as middle authors.

Decisions about authorship are not related to funding arrangements between investigators and their collaborators in the CRI Bioinformatics Unit.

In the acknowledgment, please use:

We acknowledge the support of the CRI Bioinformatics Unit, a partnership between the the Children’s Research Institute,the Center for Genetic Medicine Research, the Clinical Translational Science Institute at Childrens National (CTSI-CN) and the District of Columbia Intellectual and Developmental Disabilities Research Center (DC-IDDRC). The CTSI-CN is supported through the National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program, grant UL1TR001876 and KL2TR001877. The CTSA program is led by the NIH’s National Center for Advancing Translational Sciences (NCATS). The DC-IDDRC is supported through the National Institutes of Health (NIH) District of Columbia Intellectual and Developmental Disabilities Research Center Award (DC-IDDRC) program, grant (1U54HD090257). The DC-IDDRC program is led by NIH, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

[1] International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. http://www.icmje.org. Updated 2013. Accessed January 30, 2014.

# **Pipelines**

**RNA Seq**

 FastQ files containing the raw RNA-Seq reads output from the sequencers are first checked for quality using FastQC.

It is followed by trimming of poor quality reads using any of the following tools:

* Trimmomatic
* BBDuk
* cutadapt

An aligner is used to align the reads to the reference genome (Human/Mouse etc) using any of the following tools:

* Tophat2
* Hisat2
* STAR

The aligned Sequence Alignment Map (SAM) files are sorted and converted to BAM using samtools in some cases.

Next, the Binary Alignment Map (BAM) files are checked for their quality using either of the following tools:

* Picard RNASeqmetrics
* RSeQC

In cases for fusion gene detection, we specifically use either the Tophat-fusion or STAR-Fusion tool. While splicing events are identified using SGSe tool part of the biocondutor package in R.

Quantification of reads to Transcripts per million (TPM) is calculated using RSEM.

For cases of high sequence duplication, we tag the duplicates using Picard MarkDuplicates and count the duplicated reads using featureCounts. In many cases of overduplication we also remove the duplicated reads using Picard MarkDuplicates.

For raw unnormalized counts, we use either of the following tools:

* Htseq Counts
* featureCounts

For Differential Gene expression, we use either the following packages in R:

* Deseq2
* EdgeR
* Ebseq2

The final data can be represented in the form of heatmaps, volcano plots, MA plots and PCA plots using basic R packages.

**WGS/Exome Seq**

  The raw sequencing data in fastqc files is first quality checked with FastQC and depending upon the quality it is trimmed using the following tools.

* Trimmomatic
* BBDuk
* cutadapt

An aligner is used to align the reads to the reference genome (Human/Mouse etc) using any of the following tools:

* Bwa-mem
* Bowtie2

The aligned Sequence Alignment Map (SAM) files are processed to Binary Alignment Map (BAM) files:

* samtools
* Picard tools

We follow the Genome Analysis Toolkit 4 (GATK 4), pipeline to do Variant Calling. The BAM files are further processed for using Base Quality Score Recalibration (BQSR) step to detect and correct base calling errors from sequencers:

* BaseRecalibrator
* ApplyBQSR

Variant calling is performed on the aligned BAM files, producing genome variant call format (gvcf) files using HaplotypeCaller function.

The gvcfs are combined using CombineGVCFs function, followed by running the GenotypeGVCFs function, to increase the sensitivity of the variants called. Output of this step is a vcf file,

The vcf file is then further processed using Variant Quality Score Recalibration (VQSR) tools to increase the quality of variants identified, i.e. decrease the number of false positives.

* VariantRecalibrator
* ApplyVQSR

Annotation of variant is done using Annovar, followed by manual filtration.

**T-Cell Receptor Sequencing**

  The raw sequencing data in fastqc files is first quality checked with FastQC and depending upon the quality it is trimmed using the following tools:

* Trimmomatic
* BBDuk
* cutadapt

Extraction and alignment of fragments of target molecules is performed followed by assembly of overlapping fragmented sequencing reads into long-enough CDR3 containing contigs using the MiXCR pipeline to analyze TCR or Ig repertoire from sequencing data.

**Metagenomics/Microbiome**

 The raw sequencing data in fastqc files is first quality checked with FastQC and depending upon the quality it is trimmed using the following tools.

* Trimmomatic
* BBDuk
* cutadapt

We perform demultiplexing and quality filtering, OTU picking, taxonomic assignment, and phylogenetic reconstruction, and diversity analyses and visualizations using QIIME1 pipeline.

If you need more information about data submission or the CRI Bioinformatics Unit, please contact us at [bioinformatics@childrensnational.org](mailto:bioinformatics@childrensnational.org).

# **Workshops/Seminars**

The CNRI Bioinformatics Unit hosts seminars and workshops to Children's researchers providing them with updates on the analytical services offered by CBU.

## Past Seminars

**Topic:** [Single Cell RNA-Sequencing: A Primer](https://childrensnational.org/-/media/single-cell_rnaseq_11252019.pdf?la=en&hash=A7B36A30E403CEC1234AF3E9FCC435AA580E3770)  
**Date:**November 21, 2019

**Topic:**[Applications of Optical Genome Mapping (PDF)](https://childrensnational.org/-/media/cnhs-site/files/research-and-education/cbu/applications-of-optical-genome-mapping.pdf?la=en&hash=725F7A039D8ABCF40F506848EF58CDCC5BF4F8AC)  
**Date:**October 1, 2019

**Topic:**[Variant Identification and Interpretation from Next Generation Sequencing (PDF)](https://childrensnational.org/-/media/cnhs-site/files/research-and-education/cbu/cbu-seminar-variant-identification.pdf?la=en&hash=CF9433154A100D2EABAF3D12645830B8C08EDE7D)  
**Date:** May 28, 2019

**Topic:**[Understanding Transcriptomics with RNA Seq and Microarrays (PDF)](https://childrensnational.org/-/media/cnhs-site/files/research-and-education/cbu/cbu-seminar-transcriptomics.pdf?la=en&hash=23E0E1EDB1D1929A58592E58B49AF6AF97434060)  
**Date:**March 11, 2019

**Topic:** [Introduction to CBU and Informatics Approach to Genomics (PDF)](https://childrensnational.org/-/media/cnhs-site/files/research-and-education/cbu/cbu-seminar-introduction-to-cbu.pdf?la=en&hash=68357DFB8E2B697278C669888E9960876F00DAE5)  
**Date:** February 4, 2019