Usage Report for Fiscal Year 2020

Project Title:

Genome Immunobiology Deep Dive

Name:
○ Nicholas PARRISH (1), KOJIMA Shohei (1), Erica PARRISH (1), KAMADA Jiro (1), KOIDE Rie (1), KAWATO Satoshi (1)

(1) RIKEN Center for Integrative Medial Sciences, Genome Immunology RIKEN Hakubi Research Team

Laboratory at RIKEN:
(1) Center for Integrative Medical Sciences, Genome Immunobiology RIKEN Hakubi Research Team

1. Background and purpose of the project, relationship of the project with other projects

Animal genomes contain mobile genetic elements (MEs), such as transposons and endogenous viruses. MEs are one of the major components of animal genomes, occupying over 50% of the human and mouse genomes. Recently active MEs insert species-specific or population-specific copies in animal genomes, creating a variety of so called “structural variations” in different individual’s genomes. For example, more than 10,000 ME insertions exist in any given individual human’s genome compared to the human reference genome. This individual-to-individual variation shapes the diversity of humanity at the genomic scale, but we understand very little about how this variation influences the diversity of human form and function, or the diversity of human disease susceptibility. In large part, this is due to the complexity and enormity of the datasets which reflect this kind of genomic variation.

My team has been focusing on a particular type of ME known and endogenous viral elements, or EVEs. EVEs are animal genomic sequences that are derived from viruses. Endogenous retroviruses (ERVs) are most abundant endogenous virus in animal genomes, occupying more than 8% of the human genome. In addition to ERVs, herpesvirus and non-retroviral RNA virus-derived EVEs are also known to exist in the animal genomes. Interestingly, ERVs are known to expressed during development as well as during immune responses.

These endogenous viruses frequently code transcription regulatory sequences. Therefore, their integration proximal to genes creates novel transcriptional regulatory elements. For example, integration of endogenous retroviruses near to immune response genes has been shown to reshape inflammatory responses in mice. This exemplifies that ME insertions influence genomic and phenotypic diversity of animals and have been a driving force for animal genome evolution. However, recent integration and mobilization of these endogenous viruses are not well characterized, due to lack of analysis pipelines able to characterize them from massive genome sequence data.

This project is related to our work on endogenous viruses discovered in Japanese subjects genomes as described in Liu X et al., PLoS Genetics 2020. This project is also related to our work to uncover endogenous viral elements which were likely derived from ancient viruses or virus-like elements, described in Kojima S et al., PNAS 2021.

2. Specific usage status of the system and calculation method

We used 1,849,996.0 CPU hours and screened for mobile genetics elements in over 10,000 genomes.

3. Result

Although this was the first year of usage of HOKUSAI for this project, we discovered several previously unknown endogenous viral elements in human genomes as described in our manuscript currently under review. We also analyzed the polymorphic mobile genetic elements present in over 10,000 genomes and are currently analyzing this data. This has served as preliminary data to address the possibility that polymorphic mobile genetic elements may potentially explain antiviral immune phenotypes such as resistance to HHV-6 or severe outcome after COVID-19 infection.

4. Conclusion

Mobile genetic elements are a substantial and under-characterized component of the genetic variation between humans and other animals. Future work is needed to understand how these genetic differences influence the phenotypic differences between people.

5. Schedule and prospect for the future

Based on the benchmarking and other work performed this year in this project using HOKUSAI, we anticipate one mid-tier publication

(journal IF ~5-6) in the field of genetics. Moreover the preliminary data we generated by use of HOKUSAI has motivated and enabled us to expand the use of the bioinformatic tools developed in the course of this project to additional HPC environments on which we can analyze private human data, such as private servers of Genomics England, BioBank Japan, cloud services such as AWS, as well as alternative HPC systems in Japan designed for human genome analysis, such as SHIROKANE and the Tohoku Medical Megabank HPC environment. The speed of data transfer to HOKUSAI, especially from servers abroad such as dbGaP (NCBI) or globus which contain most of the data we access, is too slow compared to alternative systems to continue to use BW as a major computational resource for this project; this factor also explains our usage of fewer CPU hours than anticipated this year.

6. If no job was executed, specify the reason.

Fiscal Year 2020 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

<Currently under review at *PLoS Genetics*> Kojima S., Kamada, J., Parrish, N.F.* Virus-derived Variation in Diverse Human Genomes. *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.11.20.390880v1 (2020)

[Oral presentation]

“Human genome plasticity via horizontal gene transfer from human herpesvirus 6.” Society for Molecular Biology and Evolution Annual Meeting (online) June, 2020.

“Genetic variation due to viruses and other mobile genetic elements: a hidden source of variation in COVID-19-relevant phenotypes.” International Discussion Group on COVID Related Activities (organized by Mark Lathrop, McGill University). June, 2020.

“Virus-derived Structural Variation in Human Genomes: New Phenotypes from Old Viruses.” Molecular Biology Society of Japan 43rd Annual Meeting (virtual). December, 2020.

“Immunity Induced by Endogenous Viral Elements.” Japan Society for Vaccinology Annual Meeting (virtual). December, 2020.

[Poster presentation]

“Human herpesvirus 6 integrated in telomeres.” Kojima, S., Parrish, N.F. Telomere-to-Telomere Consortium and Human Pangenome Reference Consortium virtual conference, October, 2020.

[Others (Book, Press release, etc.)]

日本人ゲノムに存在する古代ウイルスの化石 ーヒトヘルペスウイルス 6 に由来する遺伝子配列の大規模解析. September 3, 2020. https://www.riken.jp/press/2020/20200903_2/index.html