BROAD INSTITUTE
CENTER FOR MENDELIAN GENOMICS

Broad CMG Publication Policy

Authorship

In general, it is anticipated that collaborators will retain first and senior authorship for discoveries made on their samples (either discovered by the core CMG staff or if made directly by the collaborators). Below are examples of different publication types and CMG authorship expectations. If you have questions about a specific publication, please feel free to reach out to Daniel MacArthur and/or Heidi Rehm for discussion of specific situations.

- **Single case in a larger series:** For publications where a CMG sequenced or analyzed sample is included in a larger case series, any CMG analyst(s) who played a major role in the discovery should be a co-author. The CMG grant must be acknowledged in the publication.

- **Cohort publications:** For publications of a cohort where the CMG provided significant sequencing and analysis support, the Broad CMG should be named as an author on resulting publications, along with the individual CMG analyst(s) who played a major role in the discovery.

- **Methods development/support:** If a discovery requires an unusually large or innovative effort made by the core CMG staff to solve certain cases (an example is an additional analysis to resolve a complex structural rearrangement), and the new method is a significant element of the paper, we expect authorship will be awarded according to the actual contributions of the participating individuals, which could involve a more prominent position for the CMG contributor.

- **CMG-led publications:** In the case where the Broad CMG leads the development of a publication (example Cummings et al. *Science Translational Medicine* Mendelian RNA-seq manuscript), investigators (and team members) who contribute to cases and analyses will be included as co-authors according to the actual contributions of the participating individuals.

Funding Citation

The CMG funding must be noted in the acknowledgments of resulting publications. Please include the following statement:

Sequencing and analysis were provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and was funded by the National Human Genome Research Institute, the National Eye Institute, and the National Heart, Lung and Blood Institute grant UM1 HG008900 and in part by National Human Genome Research Institute grant R01 HG009141.

Last updated 4/3/2020

BROAD INSTITUTE
CENTER FOR MENDELIAN GENOMICS

Broad CMG Publication Policy

NIH Public Access Policy
Any publication that received direct funding from an NIH grant (UM1 HG008900 - Joint Center for Mendelian Genomics) is required to be entered into PubMed Central within a year of publication date. As it takes a while to go through the NIHMS system, it is suggested that the submission process begin as soon as possible, and you may be marked as non-compliant if the submission process is not initiated soon after publication. You can determine the submission method for your journal via this website. If you have any questions or need any assistance with this process please contact us at cmg@broadinstitute.org.

The NIH Public Access Policy implements Division F Section 217 of PL 111-8 (Omnibus Appropriations Act, 2009). The law states:

The Director of the National Institutes of Health ("NIH") shall require in the current fiscal year and thereafter that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication: Provided, that the NIH shall implement the public access policy in a manner consistent with copyright law.

More details on which manuscripts must be in compliance

Example Methods Template Text
Sample methods text for our sequencing approaches can be seen below (if you are unsure which section is appropriate for your case, please email us):

Whole Exome Sequencing:

Long Version
Whole exome sequencing and data processing were performed by the Genomics Platform at the Broad Institute of MIT and Harvard. Libraries from DNA samples (>250 ng of DNA, at >2 ng/ul) were created with an Illumina Nextera or Twist exome capture (~38 Mb target) and sequenced (150 bp paired reads) to cover >80% of targets at 20x and a mean target coverage of >100x. Sample identity quality assurance checks were performed on each sample. The exome sequencing data was de-multiplexed and each sample's sequence data were aggregated into a single Picard BAM file.

Exome sequencing data was processed through a pipeline based on Picard, using base quality score recalibration and local realignment at known indels. The BWA aligner was used for mapping reads to the human genome build 38. Single nucleotide variants

(SNVs) and insertions/deletions (indels) were jointly called across all samples using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.5. Default filters were applied to SNV and indel calls using the GATK Variant Quality Score Recalibration (VQSR) approach. Annotation was performed using Variant Effect Predictor (VEP). Lastly, the variant call set was uploaded to *seqr* for collaborative analysis between the CMG and investigator.

Short Version
Whole exome sequencing and data processing were performed by the Genomics Platform at the Broad Institute of MIT and Harvard with an Illumina Nextera or Twist exome capture (~38 Mb target) and sequenced (150 bp paired reads) to cover >80% of targets at 20x and a mean target coverage of >100x. Exome sequencing data was processed through a pipeline based on Picard and mapping done using the BWA aligner to the human genome build 38. Variants were called using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.5.

Note about exome capture kit
CMG exome data pre-2019 was generated using an Illumina Nextera exome kit and from 2019 with a Twist exome kit.

Whole genome sequencing

Whole genome sequencing and data processing were performed by the Genomics Platform at the Broad Institute of MIT and Harvard. PCR-free preparation of sample DNA (350 ng input at >2 ng/ul) is accomplished using Illumina HiSeq X Ten v2 chemistry. Libraries are sequenced to a mean target coverage of >30x.

Genome sequencing data was processed through a pipeline based on Picard, using base quality score recalibration and local realignment at known indels. The BWA aligner was used for mapping reads to the human genome build 38. Single Nucleotide Variants (SNVs) and insertions/deletions (indels) are jointly called across all samples using the Genome Analysis Toolkit (GATK) HaplotypeCaller package version 4.0. Default filters were applied to SNV and indel calls using the GATK Variant Quality Score Recalibration (VQSR) approach. Annotation was performed using Variant Effect Predictor (VEP). Lastly, the variant call set was uploaded to *seqr* for collaborative analysis between the CMG and investigator.

Reanalysis of External Data:

Aligned whole exome or genome sequencing data was reverted back to fastq files ensuring original base quality was restored. A temporary unaligned sam file was created such that alignment independent metrics could be generated and sample QC outliers identified. The data was processed through a pipeline based on Picard, using base quality score recalibration and local realignment at known indels. We used the BWA aligner for mapping reads to the human genome build 38. Single nucleotide variants (SNVs) and insertions/deletions (indels) were jointly called across all samples using the Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.5 for exomes and 4.0 for genomes. Default filters were applied to SNV and indel calls using the GATK Variant Quality Score Recalibration (VQSR) approach. Lastly, the variants were annotated using Variant Effect Predictor (VEP). Lastly, the variant call set was uploaded to *seqr* for collaborative analysis between the CMG and investigator.

Transcriptome sequencing

Human whole transcriptome sequencing was performed by the Genomics Platform at the Broad Institute of MIT and Harvard. The transcriptome product combines poly(A)-selection of mRNA transcripts with a strand-specific cDNA library preparation, with a mean insert size of 550bp. Libraries were sequenced on the HiSeq 2500 platform to a minimum depth of 50-75 million STAR-aligned reads. ERCC RNA controls are included for all samples, allowing additional control of variability between samples.