

DATABASE ARTICLE

Open Access



mGAP: the macaque genotype and phenotype resource, a framework for accessing and interpreting macaque variant data, and identifying new models of human disease

Benjamin N. Bimber^{1,2}, Melissa Y. Yan¹, Samuel M. Peterson¹ and Betsy Ferguson^{1,3,4*}

Abstract

Background: Non-human primates (NHPs), particularly macaques, serve as critical and highly relevant pre-clinical models of human disease. The similarity in human and macaque natural disease susceptibility, along with parallel genetic risk alleles, underscores the value of macaques in the development of effective treatment strategies. Nonetheless, there are limited genomic resources available to support the exploration and discovery of macaque models of inherited disease. Notably, there are few public databases tailored to searching NHP sequence variants, and no other database making use of centralized variant calling, or providing genotype-level data and predicted pathogenic effects for each variant.

Results: The macaque Genotype And Phenotype (mGAP) resource is the first public website providing searchable, annotated macaque variant data. The mGAP resource includes a catalog of high confidence variants, derived from whole genome sequence (WGS). The current mGAP release at time of publication (1.7) contains 17,087,212 variants based on the sequence analysis of 293 rhesus macaques. A custom pipeline was developed to enable annotation of the macaque variants, leveraging human data sources that include regulatory elements (ENCODE, RegulomeDB), known disease- or phenotype-associated variants (GRASP), predicted impact (SIFT, PolyPhen2), and sequence conservation (PhyloP, PhastCons). Currently mGAP includes 2767 variants that are identical to alleles listed in the human ClinVar database, of which 276 variants, spanning 258 genes, are identified as pathogenic. An additional 12,472 variants are predicted as high impact (SnPEff) and 13,129 are predicted as damaging (PolyPhen2). In total, these variants are predicted to be associated with more than 2000 human disease or phenotype entries reported in OMIM (Online Mendelian Inheritance in Man). Importantly, mGAP also provides genotype-level data for all subjects, allowing identification of specific individuals harboring alleles of interest.

Conclusions: The mGAP resource provides variant and genotype data from hundreds of rhesus macaques, processed in a consistent manner across all subjects (<https://mgap.ohsu.edu>). Together with the extensive variant annotations, mGAP presents unprecedented opportunity to investigate potential genetic associations with currently characterized disease models, and to uncover new macaque models based on parallels with human risk alleles.

Keywords: SNP, *Macaca mulatta*, Nonhuman primate, Genome, Rhesus, Indian-origin, Animal model, Database

* Correspondence: fergusob@ohsu.edu

¹Division of Genetics, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA

³Division of Neuroscience, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA

Full list of author information is available at the end of the article



Background

One of the most exciting outcomes from the genomic era and the plunging costs of DNA sequencing is the increased efficiency to identify genetic variants associated with disease risk and progression. This leap forward portends the rapidly approaching era of predictive genetic medicine and personalized therapy. Translating insight into the clinic will require robust genomic models to test and advance gene-based therapeutic approaches. Non-human primates (NHPs), especially macaques, are well positioned to serve as a critical link, providing a highly relevant pre-clinical model. Due to their high level of genomic, physiological, anatomical, neurological and behavioral similarity to humans, macaques are often indispensable to the study of complex human disease.

The close evolutionary history of macaques and humans is evident in their highly similar breadth of natural disease susceptibilities. Moreover, the parallel genetic associations already reported for a broad range of diseases (Table 1) suggests opportunity to leverage the macaque for the discovery of disease-associated biomarkers and for use in the development of new pharmacogenomic or personalized medicine therapies [1–30]. The range of orthologous genetic associations include

common diseases, such as cancers, reproductive disorders, retinal and infectious diseases, and rare maladies such as Krabbe disease. Similarly, complex behavioral traits such as heightened anxiety, and variable response to commonly used pharmacological agents such as naltrexone, have been linked to the variants in the same genes as in humans. Other macaque disease models, such as for autism, polycystic ovarian syndrome (PCOS) or type II diabetes, have been reported, but genetic linkages have yet to be uncovered (Table 1). The growing list of macaque traits are likely the ‘tip of the iceberg’ of the natural models of human disease, as other diseases likely go undetected either due to a less obvious phenotype (e.g., hearing loss) or due to early life lethality. The broad spectrum of natural models in rhesus macaque populations presents largely untapped opportunities for translational study, and may negate the need for more costly model development approaches, such as gene-editing.

Despite the clear value of macaques as a human disease models, there are very limited genomic resources available to support the exploration and discovery of potential natural models of genetic disease. Notably, there is a paucity of public databases tailored to searching NHP sequence variant and phenotypic data. The major human variant databases, dbSNP [31] and dbVar [32], phased out support for non-human data in 2017. The NHP Research Consortium Genetic Variant Database website provides a portal for users to deposit NHP SNP data [33]. While valuable, related information about allele frequencies, predicted functional effects, cross-species conservation or links to source animals are not available, limiting the utility for variant interpretation. The Ensembl databases include macaque data, which includes a diverse set of databases and tools, a genome browser, and varying degrees of population-level and phylogenetic data. Here we present the macaque Genotype And Phenotype (mGAP) resource, the first public website providing searchable, extensively annotated macaque genomic data generated from a large, well-characterized rhesus macaque cohort. Once registered, users can browse a curated catalog of variants derived from the whole genome sequence analysis of a large and growing cohort of rhesus macaques. The website and dataset complement existing resources such as Ensembl, primarily by providing data from a defined cohort, much like the 1000Genomes project or similar human cohorts [34]. Importantly, because mGAP provides genotype-level data for all subjects, it is possible to identify specific individuals harboring alleles of interest. We developed a novel annotation pipeline that leverages human annotation data sources to guide interpretation of the macaque variants, an approach that could easily be adapted for other model organisms. The website is designed to allow direct exploration of genomic data, with easy export of all information in standard conventional formats for external processing.

Table 1 Example of diseases and traits reported in rhesus macaques. Diseases are grouped based on whether a genetic association has been reported in macaques for that disease

Genetic Association Reported
SIV progression [1–3]
Age-related macular degeneration [4, 5]
Colorectal cancer [6]
Krabbe disease [7]
Ammenorhea [8]
Naltrexone response [9]
HPA axis dysregulation [10]
Alcohol intake levels [11–13]
Anxiety [14, 15]
No Genetic Association Reported
Autism [16]
Age-associated cognitive decline [17]
Polycystic ovary syndrome (PCOS) [18]
Endometriosis [19–21]
Left ventricular hypertrophy [22]
Type III Von-Willebrand [23]
Type II Diabetes [24–26]
Squamous cell carcinoma [27]
Coat-like retinopathy [28]
TB progression [29]
Osteoporosis [30]

mGAP represents an important new resource for NHP genetic and genomic research, which together with the expanding range of macaque ‘omics’ data, supports investigation into links between sequence variants and molecular mechanisms of disease.

Construction and content

Rhesus macaque variant catalog

A key feature of mGAP is a catalog of high confidence genomic variants, derived from high depth, whole genome data. The current mGAP release at the time of this publication (1.7) contains whole genome data from 293 Indian-origin rhesus macaques. These animals were selected from a multi-generational outbred cohort of macaques bred at the Oregon National Primate Research Center (average kinship 0.0016). Genomic DNA was isolated, used to generate Illumina libraries as previously described, followed by sequencing on an Illumina XTen sequencer [35]. We obtained an average of 704,193,015 paired Illumina reads per animal, with a range of 549,660,386 - 1,843,458,162 (Additional file 1: Table S1). Primary FASTQ files have been submitted to SRA under BioProject PRJNA382404. All data were aligned to the MMuL_8.0.1 genome (GCF_000772875.2), with an average of 700,505,668 paired reads aligned per animal, providing an average high-confidence genome coverage of 20X. These sequence data were processed using a previously described macaque variant calling pipeline ([35] and Additional file 2: Supplemental Methods), adapted from the Broad Institute Best Practice Variant Discovery Pipeline [36]. The pipeline includes extensive quality filtering of the raw variant data as previously described, including the use of Mendelian inheritance for validation.

The mGAP 1.7 release contains 17,087,212 passing variants from this dataset (Fig. 1a). Of these, 2,664,020 (15.5%) are private alleles (detected in only a single individual). Because the cohort includes distantly related individuals, the latter is likely lower than would be observed in an unrelated cohort of similar size. These data include single nucleotide variants (SNVs) and short indels, with the latter comprising 12.0% of variants. As expected, the majority of variants are outside of coding regions (Fig. 1b); however, we identified 250,030 variants within coding regions, including 4888 frameshift mutations and 1871 stop-gained mutations. The mGAP variant database is updated approximately 3–4 times per year, as new data become available.

Variant functional annotation

Assigning functional significance to genetic variants is often a challenge, and this is particularly problematic for model organisms like macaques that have far less developed annotation resources than human. To overcome this, we developed a multi-part process to annotate our

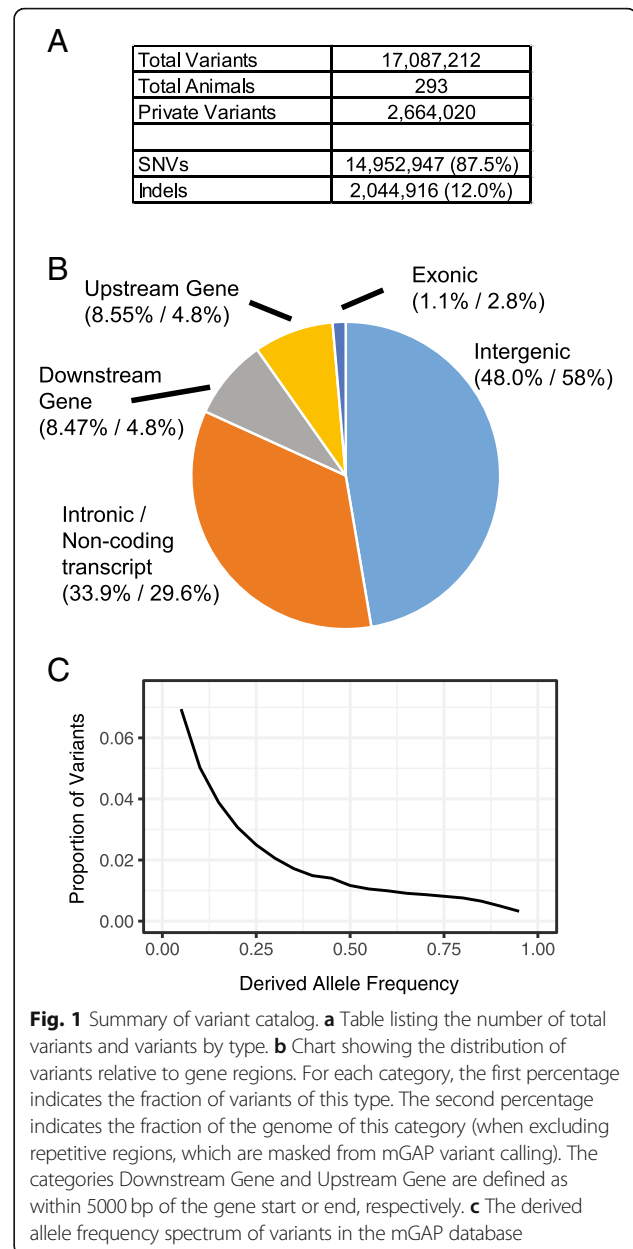
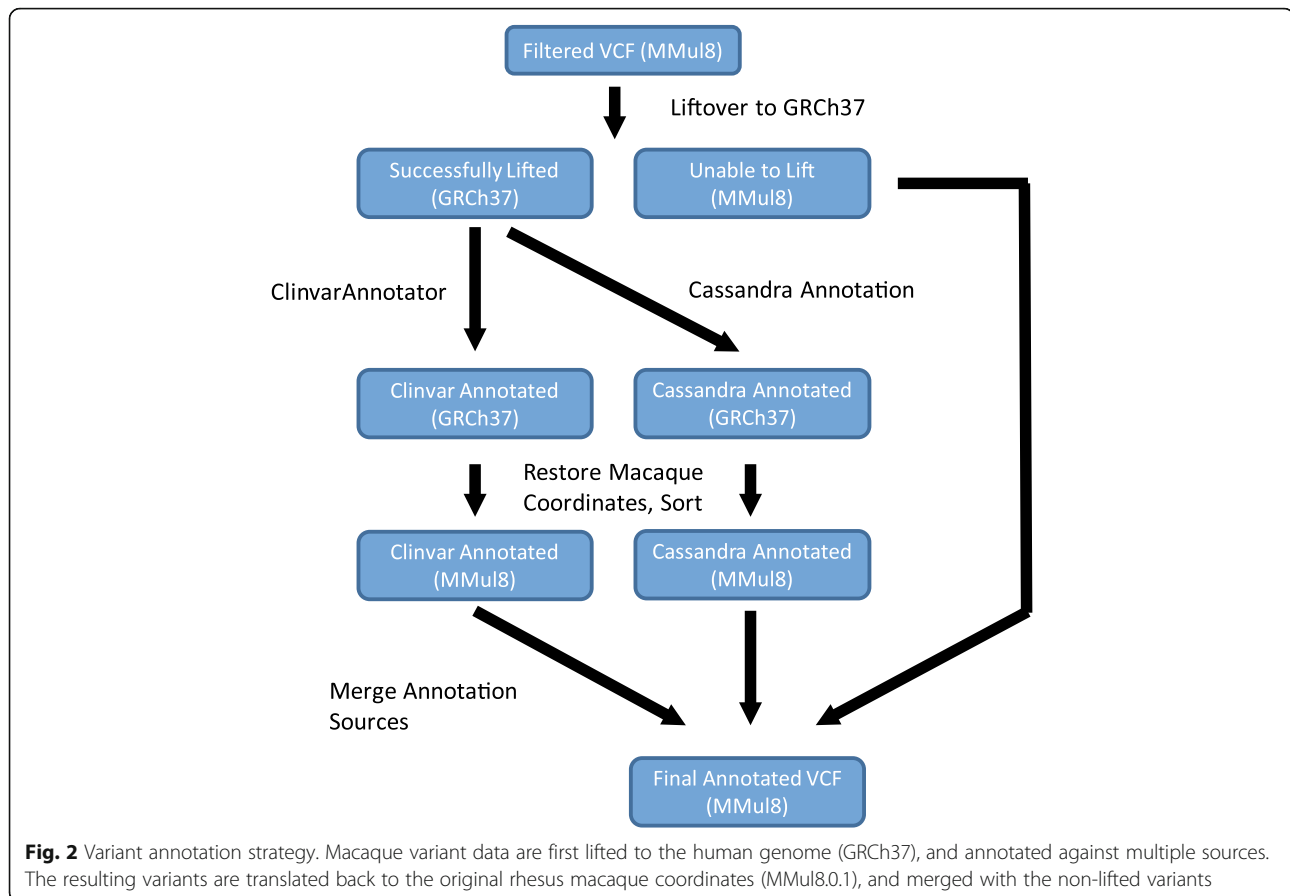


Fig. 1 Summary of variant catalog. **a** Table listing the number of total variants and variants by type. **b** Chart showing the distribution of variants relative to gene regions. For each category, the first percentage indicates the fraction of variants of this type. The second percentage indicates the fraction of the genome of this category (when excluding repetitive regions, which are masked from mGAP variant calling). The categories Downstream Gene and Upstream Gene are defined as within 5000 bp of the gene start or end, respectively. **c** The derived allele frequency spectrum of variants in the mGAP database

data (Fig. 2). Rhesus macaque variants are first lifted to the human genome (GRCh37.p13, GCA_000001405.14). In the latest release, approximately 65% of macaque variants were successfully lifted to the human genome. Any variant that was unable to lift to the human genome was annotated accordingly. Of the variants that lifted to human, 3,046,012 (17.8%) matched positions reported in the 1000Genomes phase 3 dataset, representing a range of macaque allele frequencies (Additional file 3: Figure S1). The successfully lifted variants were annotated using multiple sources. We utilized Cassandra, a utility that annotates variants overlapping with a variety of human data sources [37], including regulatory elements



(ENCODE [38], RegulomeDB [39]), overlap with known disease- or phenotype-associated variants (GRASP [40]), predicted impact (CADD [41, 42], SIFT [43], PolyPhen2 [44]), conservation (PhyloP [45], PhastCons [45]), and others [37, 46, 47]. We additionally annotated any variants with identical position and allele as a variant recorded in ClinVar as ‘pathogenic’ or ‘likely pathogenic’ (ClinVar download 9/1/2018). [48–50]. After annotation, variants were translated back to their original macaque genomic coordinates. In addition to human annotations, SnpEff was used to annotate predicted impact on protein coding using the macaque gene annotations [51]. While developed for rhesus macaque, the same liftover annotation process could easily be adapted to other model organisms. The final product of the annotation pipeline is a single annotated VCF file, which is a standard and interchangeable format. All tools created for this process are publicly available [52].

Using this annotation pipeline, there were 2767 variants identical at both position and allele to variants in the human ClinVar database (version GRCh37 20,180,128). Of these, 276 are listed as pathogenic, spanning 258 distinct genes. This list includes variants linked to 160 distinct disorders/phenotypes, including cardiovascular phenotypes (aortic aneurysm, congenital heart disease), immune

disorders (polyglandular autoimmune syndrome, Behcet’s syndrome), and many additional rare diseases (Peutz-Jeghers syndrome, Leigh syndrome). Our annotation process also identified many variants predicted to severely impact protein coding, including 12,472 predicted as high impact by SnpEff and 13,129 predicted as damaging by PolyPhen2, encompassing 4091 genes with predicted frameshift mutations. A complete list of variants is available through the mGAP website.

mGAP website design and data storage

The mGAP website was developed as a module for LabKey Server [53], written using a mixture of Java and JavaScript. The site was designed to heavily leverage existing public resources, such as integration with JBrowse, an open-source genome browser [54]. The source code for the mGAP module and all related modules is freely available through the LabKey Server public subversion repository (<https://svn.mgt.labkey.host/stedi/trunk>). The primary mode of data storage for each release is the annotated VCF file, which has the advantage of being easily exportable and capable of being processed with most standard variant-processing tools. Limited summary statistics on each release are stored in the database, such as the number

of variants and a breakdown of variants by type. The site includes demographics data on all animals, which is also stored and implemented using the LabKey study framework [53].

Utility and discussion

The mGAP website overview

The mGAP website provides an interface to download raw data or to explore macaque variants, available at <https://mgap.ohsu.edu/>. The main page of the website contains a summary of the current release, and tabs for each of the main datatypes: the variant catalog, macaque phenotype and model information, cohort/demographics data and raw sequence data (Fig. 3). The mGAP database currently includes a list of disease models and phenotypes reported in rhesus macaques. The website provides a table with demographics data from the cohort, including sex, species, age, ancestry, and average kinship within the cohort. The mGAP website will be expanded in the future to include clinical data and other phenotypic measures for each subject as well.

Access to variant data

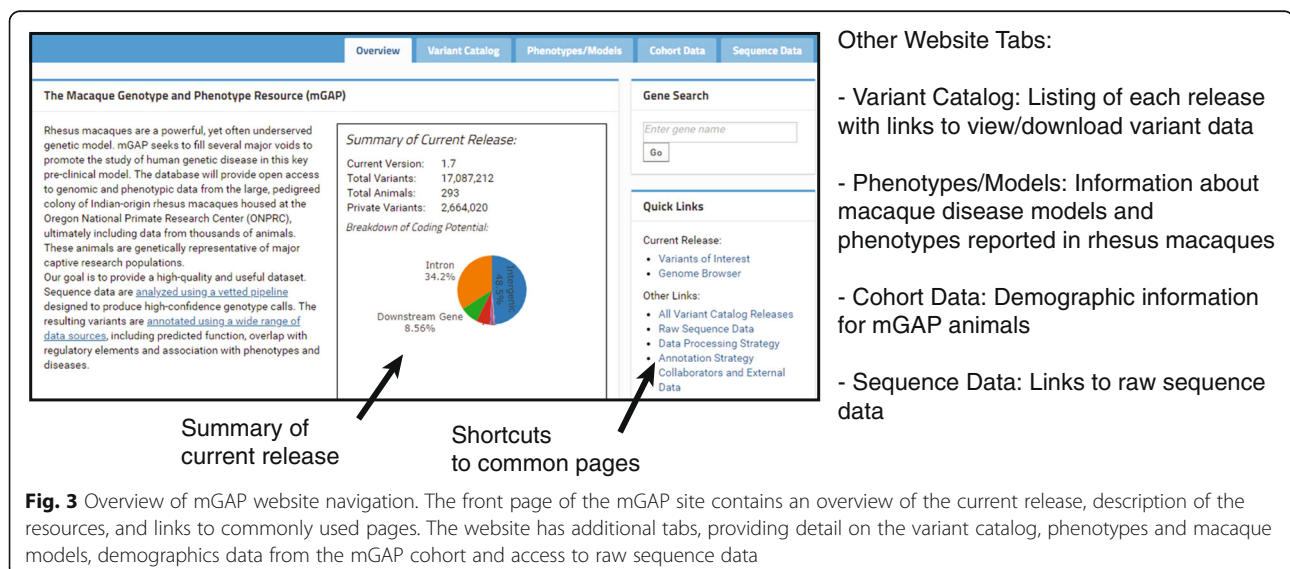
The website is designed to allow any user to easily interact with variant data and to support basic queries of the dataset. The genome browser is the primary mode through which users can interactively explore variants. Users can search by genomic position or gene name, and view variants in the context of genes and other macaque genomic features. Variants are color-coded by type and functional prediction to highlight predicted missense and high-impact variants (Fig. 4). Users can click on any variant to view greater detail, including the complete set of functional annotations and cross-species conservation data. The variant details window also includes the minor

allele frequency within the currently sequenced population, distribution of homozygous and heterozygous genotypes, and a link to view individual genotypes at this position (as a separate table). In addition, for each release we generate a list of predicted high-impact variants, intended to include those variants either previously implicated in human disease (such as those overlapping ClinVar data), or predicted to have a high impact on protein coding (such as SnpEff high-impact or PolyPhen2 damaging variants). In the 1.7 release, this high-impact list comprises 25,877 genome-wide variants that are predicted to be associated with 2071 human disease or phenotype entries listed in OMIM (Online Mendelian Inheritance in Man).

Conclusions

The macaque Genotype and Phenotype Resource (mGAP) represents a critical resource for genomic research that fills a major gap in the NHP research field. mGAP provides a unique combination of variant and genotype data from a large cohort of Indian-origin rhesus macaques, processed in a consistent manner across all subjects. While comparable datasets are common for most models, until now they have been lacking for macaques. These data serve basic yet essential functions to support modern genomic analyses, including a catalog of common and rare variants in macaques, coupled with allele frequency and genotype-level data. The mGAP website is designed to allow exploration of data, primarily through the genome browser, or complete export of the entire dataset into common formats for external analysis.

The resources provided by mGAP are designed to advance genomic research in macaques, with the primary goal of identifying and developing new genetic disease models. In this regard, the resource has utility that



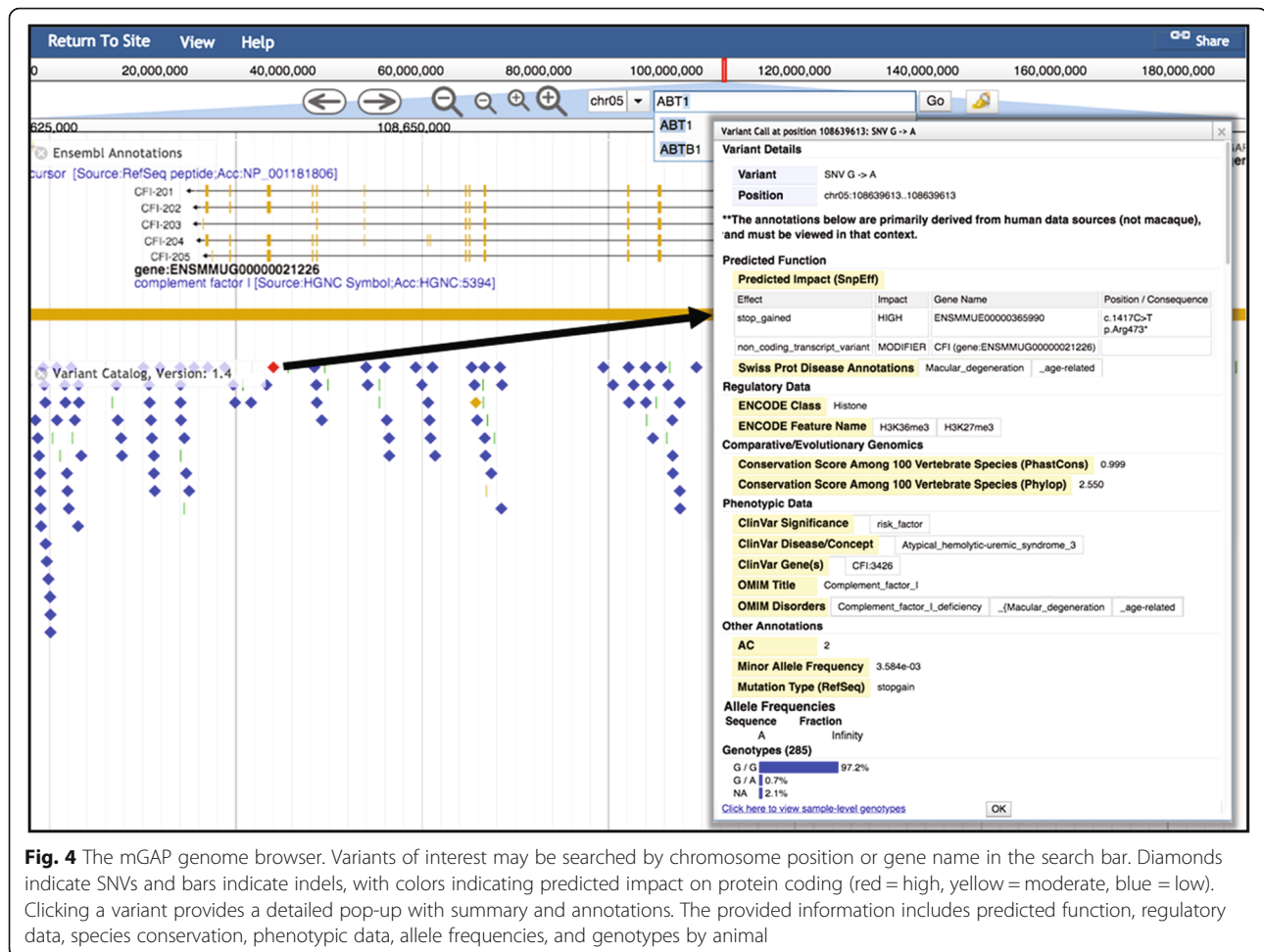


Fig. 4 The mGAP genome browser. Variants of interest may be searched by chromosome position or gene name in the search bar. Diamonds indicate SNVs and bars indicate indels, with colors indicating predicted impact on protein coding (red = high, yellow = moderate, blue = low). Clicking a variant provides a detailed pop-up with summary and annotations. The provided information includes predicted function, regulatory data, species conservation, phenotypic data, allele frequencies, and genotypes by animal

extends beyond the initial set of sequenced animals. Investigators can explore potential associations between a macaque phenotype of interest and variants in candidate genes by using the mGAP browser to identify compelling candidate variants based on predicted translational effects, allele frequencies and cross species conservation. Prioritized variants can then be evaluated using targeted re-sequencing or genotype-specific assays in macaques lacking genome-wide sequence and tested for phenotype association.

mGAP also presents new opportunities to reveal currently unrecognized macaque disease models, which is facilitated by the highlighted ‘Variants of Interest’ list on the website. This list includes variants that are either identical to those linked to disease in humans or have a predicted highly deleterious impact on protein coding. Supporting this ‘reverse genetic’ approach to model discovery, a comprehensive medical record is established for each of the rhesus macaques housed at the Oregon National Primate Research Center, including regularly updated clinical data, and as available, imaging, histology, and pathology notations. The identification of

naturally occurring genetic macaque models provides several advantages over genetically-induced models, such as those produced using CRISPR-like approaches, since there is no danger of experimentally induced off-target effects, the phenotypic evaluation can be carried out in existing animals at considerably less cost than creation of transgenic animals, and the identification of multiple subjects with the desired genotype has the potential to support model expansion.

The rapid expansion of ‘-omic’ level data produced by both NHP and human studies, affords exciting new opportunities for high-resolution data integration and analysis. In this regard, mGAP provides a missing link in the coordinated study of genomic variants in combination with macaque transcriptomic, metabolomic and proteomic findings [55–58]. In parallel, deep learning framework approaches are now emerging that enable the prediction of the functional effects of single base variants on gene expression and disease risk [59]. In addition to purely macaque datasets, studies have shown the value of combined analysis of human and NHP genome-wide data, through the use of neural network

approaches that predict variant association with disease, and mGAP data may facilitate similar studies [60].

In addition to model discovery and analysis, mGAP also provides a much-needed resource to improve the design of molecular research tools. For example, information on variant location and allele frequency is critical information to optimize the design of DNA primers for use in PCR amplification or peptide-based ligands for antibody generation. Similarly, SNP array designs for research use or colony management applications will benefit from the available information on allele frequencies and potential pathogenic effects, guiding the selection of appropriate variants to meet the goals of each tool.

Interpretation of genome-wide data can be daunting. For human genomic data, tools continue to be developed to sift through variants and prioritize potentially damaging or relevant variants. The algorithms used in these tools are generally aided by the considerable body of human genome annotation, ranging from catalogs of disease-associated variants to functional annotation developed by experimental studies. Because these resources do not currently exist for macaques, we developed a process to annotate macaque variants using available human data. The annotation processes developed for mGAP can also serve as a template, adapted to enhance the interpretation of genomic variants associated with other model organisms. While lifting human annotation to macaque has value and can identify relevant variants, interpreting annotation across species must be viewed carefully and can have pitfalls, especially for non-coding and regulatory regions. It is therefore important to develop macaque-specific functional annotation, such as data linking genetic variation to changes in gene expression (i.e. eQTLs) or other functional genomic data. The rapid decrease in sequence costs, and rise of massively parallel single-cell analysis approaches, should enable these data to be generated much more rapidly and at a fraction of the cost of the analogous human studies.

While the majority of mGAP data is centered on Indian-origin rhesus macaques from the ONPRC breeding colony, previous studies have found that these macaques are genetically representative of major research captive breeding populations at other US National Primate Research Centers [61]. Further, the utility of mGAP has attracted interest from a broad range of NHP investigators, and thus mGAP now supports data submissions from other rhesus macaque populations. This expansion in data sources will enable immediate cross referencing of variants of interest, facilitating the identification of genotype-specific animals at multiple facilities. In the advancing era of precision medicine, selection and access to genotyped animals will be

increasingly critical to enabling state-of-the-art biomedical research.

Additional files

Additional file 1: Table S1. Summary table of per-sample sequence data. (XLSX 27 kb)

Additional file 2: Supplemental Methods. (DOCX 24 kb)

Additional file 3: Figure S1. Derived allele frequency spectrum of macaque variants shared with the 1000Genomes Phase 3 dataset. The proportion of variants for each allele frequency is shown for the entire mGAP release 1.7 dataset (blue), and the subset of variants that intersect with sites reported in the 1000Genomes Phase 3 dataset (red). (PDF 69 kb)

Abbreviations

mGAP: macaque Genotype And Phenotype; NHP: Non-human primate; OMIM: Online Mendelian Inheritance in Man; PCOS: Polycystic ovarian syndrome; WGS: Whole genome sequence

Acknowledgements

Not applicable

Funding

This work was funded by NIH grants P51OD011092, and R24OD021324. The funding bodies had no direct role in the design, data collection, analysis, data interpretation, or writing of the manuscript.

Availability of data and materials

The raw sequence data used in mGAP are available in SRA under BioProject PRJNA382404. The source code for the mGAP LabKey modules and related modules are open-source and available in the LabKey subversion repository (<https://svn.mgt.labkey.host/stedi/trunk>). The current and historic mGAP releases are available as VCF files for download via the mGAP website (<https://mgap.ohsu.edu/>).

Authors' contributions

BNB participated in data collection, analysis, as well as the design of analysis and website code. MYY contributed to data collection, design of analysis code and data analysis. SMP participated in data collection and analysis of variant data. BF oversaw the project, and contributed to data analysis and interpretation. All authors contributed to manuscript writing and review. All authors read and approved the final manuscript.

Ethics approval

All sample collection protocols were approved by the Oregon Health & Sciences University Animal Utilization and Care Committee and in accordance with the NIH and the Guide for Use and Care of Laboratory Animals.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Division of Genetics, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA. ²Division of Pathobiology, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA. ³Division of Neuroscience, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA. ⁴Molecular and Medical Genetics Department, Oregon Health and Sciences University, Portland, OR 97239, USA.

Received: 20 October 2018 Accepted: 22 February 2019

Published online: 06 March 2019

References

- Nomura T, Matano T. Association of MHC-I genotypes with disease progression in HIV/SIV infections. *Front Microbiol.* 2012;3:234.
- Letvin NL, et al. Immune and Genetic Correlates of Vaccine Protection Against Mucosal Infection by SIV in Monkeys. *Sci Transl Med.* 2011;3(81):81ra36.
- Lim SY, et al. Contributions of Mamu-a*01 status and TRIM5 allele expression, but not CCL3L copy number variation, to the control of SIVmac251 replication in Indian-origin rhesus monkeys. *PLoS Genet.* 2010; 6(6):e1000997.
- Francis PJ, et al. Rhesus monkeys and humans share common susceptibility genes for age-related macular disease. *Hum Mol Genet.* 2008;17(17):2673–80.
- Singh KK, et al. Association of HTRA1 and ARMS2 gene variation with drusen formation in rhesus macaques. *Exp Eye Res.* 2009;88(3):479–82.
- Brammer DW, et al. MLH1-rhMac hereditary nonpolyposis colorectal cancer syndrome in rhesus macaques. *Proc Natl Acad Sci U S A.* 2018; 115(11):2806–11.
- Luzi P, et al. Characterization of the rhesus monkey galactocerebrosidase (GALC) cDNA and gene and identification of the mutation causing globoid cell leukodystrophy (Krabbe disease) in this primate. *Genomics.* 1997;42(2):319–24.
- Lomniczi A, et al. A single-nucleotide polymorphism in the EAP1 gene is associated with amenorrhea/oligomenorrhea in nonhuman primates. *Endocrinology.* 2012;153(1):339–49.
- Vallender EJ, et al. A pharmacogenetic model of naltrexone-induced attenuation of alcohol consumption in rhesus monkeys. *Drug Alcohol Depend.* 2010;109(1–3):252–6.
- Ferguson B, et al. Genetic load is associated with hypothalamic-pituitary-adrenal axis dysregulation in macaques. *Genes Brain Behav.* 2012;11(8):949–57.
- Lindell SG, et al. Functional NPY variation as a factor in stress resilience and alcohol consumption in rhesus macaques. *Arch Gen Psychiatry.* 2010;67(4):423–31.
- Barr CS, et al. Functional CRH variation increases stress-induced alcohol consumption in primates. *Proc Natl Acad Sci U S A.* 2009;106(34):14593–8.
- Cervera-Juanes R, et al. MAOA expression predicts vulnerability for alcohol use. *Mol Psychiatry.* 2016;21(4):472–9.
- Spinelli S, et al. The serotonin transporter gene linked polymorphic region is associated with the behavioral response to repeated stress exposure in infant rhesus macaques. *Dev Psychopathol.* 2012;24(1):157–65.
- Rogers J, et al. CRHR1 genotypes, neural circuits and the diathesis for anxiety and depression. *Mol Psychiatry.* 2013;18(6):700–7.
- Bauman MD, Schumann CM. Advances in nonhuman primate models of autism: Integrating neuroscience and behavior. *Exp Neurol.* 2018;299(Pt A):252–65.
- Paspalas CD, et al. The aged rhesus macaque manifests Braak stage III/IV Alzheimer's-like pathology. *Alzheimers Dement.* 2018;14(5):680–91.
- Abbott DH, et al. Clustering of PCOS-like traits in naturally hyperandrogenic female rhesus monkeys. *Hum Reprod.* 2017;32(4):923–36.
- Assaf BT, Miller AD. Pleural endometriosis in an aged rhesus macaque (*Macaca mulatta*): a histopathologic and immunohistochemical study. *Vet Pathol.* 2012;49(4):636–41.
- Franasiak JM, et al. Endometrial CXCL13 expression is cycle regulated in humans and aberrantly expressed in humans and rhesus macaques with endometriosis. *Reprod Sci.* 2015;22(4):442–51.
- Zondervan KT, et al. Familial aggregation of endometriosis in a large pedigree of rhesus macaques. *Hum Reprod.* 2004;19(2):448–55.
- Reader JR, et al. Left ventricular hypertrophy in rhesus macaques (*Macaca mulatta*) at the California National Primate Research Center (1992–2014). *Comp Med.* 2016;66(2):162–9.
- Patterson MM, et al. Type-3 von willebrand's disease in a rhesus monkey (*Macaca mulatta*). *Comp Med.* 2002;52(4):368–71.
- Qian C, et al. Diastolic dysfunction in spontaneous type 2 diabetes rhesus monkeys: a study using echocardiography and magnetic resonance imaging. *BMC Cardiovasc Disord.* 2015;15:59.
- Pare M, et al. Differential hypertrophy and atrophy among all types of cutaneous innervation in the glabrous skin of the monkey hand during aging and naturally occurring type 2 diabetes. *J Comp Neurol.* 2007; 501(4):543–67.
- Bremer AA, et al. Fructose-fed rhesus monkeys: a nonhuman primate model of insulin resistance, metabolic syndrome, and type 2 diabetes. *Clin Transl Sci.* 2011;4(4):243–52.
- Jean SM, et al. Spontaneous primary squamous cell carcinoma of the lung in a rhesus macaque (*Macaca mulatta*). *J Am Assoc Lab Anim Sci.* 2011; 50(3):404–8.
- Liu DX, et al. Coats-like retinopathy in a young Indian rhesus macaque (*Macaca mulatta*). *J Med Primatol.* 2015;44(2):108–12.
- Maiello P, et al. Rhesus Macaques are more susceptible to progressive tuberculosis than *Cynomolgus* Macaques: a quantitative comparison. *Infect Immun.* 2018;86(2).
- Simmons HA. Age-associated pathology in rhesus macaques (*Macaca mulatta*). *Vet Pathol.* 2016;53(2):399–416.
- Sherry ST, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308–11.
- Lappalainen I, et al. DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res.* 2013;41(Database issue):D936–41.
- Nonhuman Primate Genetic Variant Database. Available from: <https://nprcresearch.org/primate/genetics-genomics/nonhuman-primate-genetic-variant-database.php>.
- Genomes Project, C, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
- Bimber BN, et al. Whole genome sequencing predicts novel human disease models in rhesus macaques. *Genomics.* 2017;109(3–4):214–20.
- Van der Auwera GA, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:11 10 1–33.
- Cassandra: A tool for annotating genomic variant data. Available from: <https://www.hgsc.bcm.edu/software/cassandra>.
- Consortium EP. The ENCODE (ENCyclopedia of DNA elements) Project. *Science.* 2004;306(5696):636–40.
- Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790–7.
- Leslie R, O'Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics.* 2014;30(12):i185–94.
- Rentsch P, et al. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886–94.
- Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310–5.
- Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812–4.
- Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013; Chapter 7:Unit7 20.
- Ramani R, et al. PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. *Bioinformatics.* 2018.
- Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
- Smigielski EM, et al. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.* 2000;28(1):352–5.
- Landrum MJ, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062–7.
- Landrum MJ, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014; 42(Database issue):D980–5.
- Clinvar 2.0 VCF File. Available from: ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/archive_2.0/2018/clinvar_20180128.vcf.gz.
- Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012; 6(2):80–92.
- DISCVr-Seq. DISCVr-Seq: A set of command line tools for working with sequence data. Available from: <https://github.com/BimberLab/DISCVrSeq>.
- Nelson EK, et al. LabKey server: an open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics.* 2011;12:71.
- Buels R, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 2016;17:66.
- Lee KJ, et al. Comparative transcriptomics and metabolomics in a rhesus macaque drug administration study. *Front Cell Dev Biol.* 2014;2:54.

56. MacDonald ML, et al. Altered glutamate protein co-expression network topology linked to spine loss in the auditory cortex of schizophrenia. *Biol Psychiatry*. 2015;77(11):959–68.
57. Yu X, et al. Quantitative proteomics reveals the novel co-expression signatures in early brain development for prognosis of glioblastoma multiforme. *Oncotarget*. 2016;7(12):14161–71.
58. Bakken TE, et al. A comprehensive transcriptional map of primate brain development. *Nature*. 2016;535(7612):367–75.
59. Zhou J, et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat Genet*. 2018;50(8):1171–9.
60. Sundaram L, et al. Predicting the clinical impact of human mutation with deep neural networks. *Nat Genet*. 2018;50(8):1161–70.
61. Kanthaswamy S, et al. Development and validation of a SNP-based assay for inferring the genetic ancestry of rhesus macaques (*Macaca mulatta*). *Am J Primatol*. 2014;76(11):1105–13.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

