SOFTWARE   **Open Access**

MySeq: privacy-protecting browser-based personal Genome analysis for genomics education and exploration

Michael D. Linderman* , Leo McElroy and Laura Chang

Abstract

Background: The complexity of genome informatics is a recurring challenge for genome exploration and analysis by students and other non-experts. This complexity creates a barrier to wider implementation of experiential genomics education, even in settings with substantial computational resources and expertise. Reducing the need for specialized software tools will increase access to hands-on genomics pedagogy.

Results: MySeq is a React.js single-page web application for privacy-protecting interactive personal genome analysis. All analyses are performed entirely in the user's web browser eliminating the need to install and use specialized software tools or to upload sensitive data to an external web service. MySeq leverages Tabix-indexing to efficiently query whole genome-scale variant call format (VCF) files stored locally or available remotely via HTTP(s) without loading the entire file. MySeq currently implements variant querying and annotation, physical trait prediction, pharmacogenomic, polygenic disease risk and ancestry analyses to provide representative pedagogical examples; and can be readily extended with new analysis or visualization components.

Conclusions: MySeq supports multiple pedagogical approaches including independent exploration and interactive online tutorials. MySeq has been successfully employed in an undergraduate human genome analysis course where it reduced the barriers-to-entry for hands-on human genome analysis.

Keywords: Personal Genome analysis, Genomics education, Web application

Background

The growing deployment of genome sequencing in research, clinical and commercial contexts is creating a corresponding need for more effective and scalable genomics pedagogy for both providers and patients/participants [1–10]. New genomics curricula are in development to provide students hands-on experience tackling the increased scale and complexity of genome sequencing data [11–19]. However the complexity of genome informatics is a recurring challenge, even in settings with substantial computational resources and expertise [20, 21], creating a barrier to wider implementation of experiential genomics education [22]. Reducing the need for command-line and other specialized software will increase student access to hands-on genome analysis experiences.

Web applications can provide an easier-to-use alternative to command-line and other specialized software. In a traditional "server-side" web application the genomic analyses would be performed on a remote server. Modern web technologies, however, enable genomic analyses to be performed entirely in the user's web browser. This "client-side" approach can provide the same ease-of-use while protecting the privacy of users' sensitive genomic data (no data is uploaded to a remote server) and minimizing the infrastructure required for hands-on genomic analysis (no need for an application server). Ensuring users maintain control over their genomic data is a particularly important feature for the growing number of courses in which students analyze their own genomic data [11, 23–27].

GENOtation (formerly named the Interpretome) [28] is a web browser-based genome interpretation tool developed to support students' analysis of their microarray genotyping data [26]. GENOtation loads the genotyping

* Correspondence: mlinderman@middlebury.edu
Department of Computer Science, Middlebury College, Middlebury, VT, USA

data locally from the user's computer and performs the analyses exclusively within the browser. GENOtation is not designed, however, for use with variant call format (VCF) files commonly produced by whole exome and genome sequencing (WES/WGS). DNA Compass [29] employs a similar browser-based model for querying locally-stored VCF files downloaded from the DNA.Land digital biobank [30] (or other sources) and linking those variants to public databases, but does not implement other analyses. The iobio suite [31, 32] includes applications for combined browser and server-based analysis of locally-stored or remotely-available VCF files but is focused on filtering for putative disease variants. Web-based genome browsers and pileup viewers, such as the UCSC Genome Browser [33], JBrowse [34], igv.js [35] and pileup.js [36], can display remotely-available coordinate-indexed VCF files without additional software and some tools can also display locally-stored VCF files (e.g., igv.js and JBrowse), but a genome browser only provides limited variant analysis functionality (primarily query by genomic region).
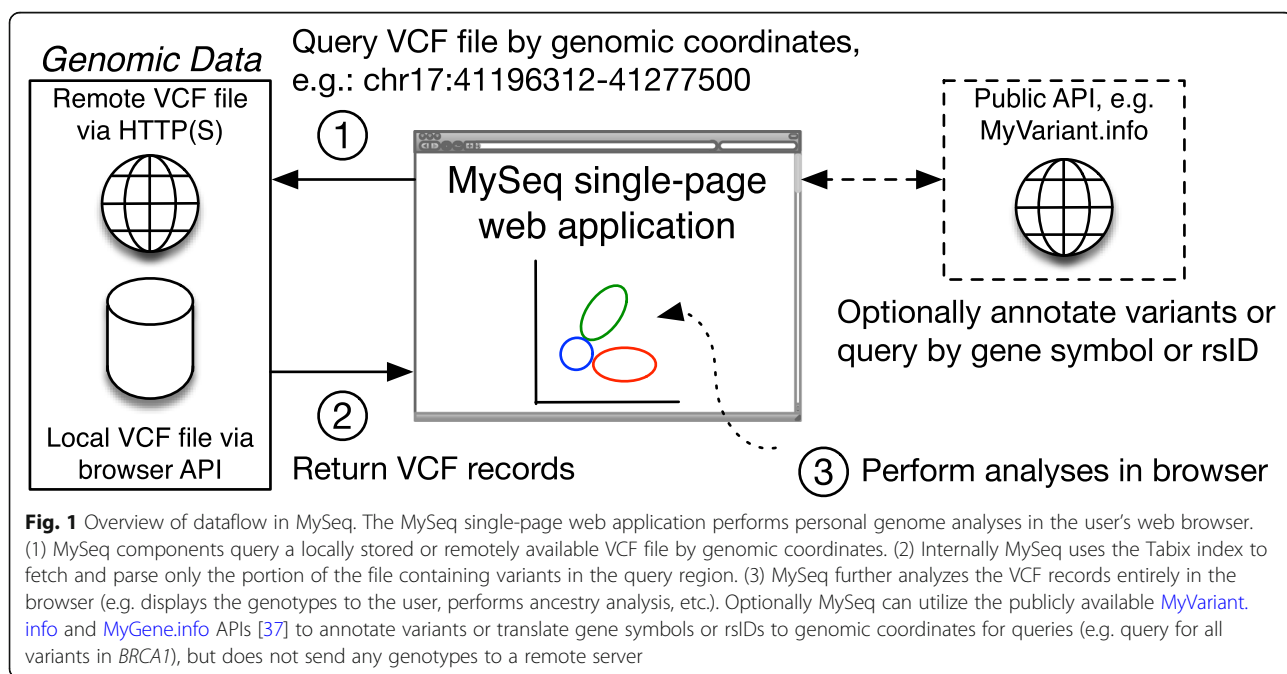
Here we present MySeq, a freely available open-source web application, inspired by GENOtation, DNA Compass and the iobio suite, which is designed to meet the unique needs of experiential genomics pedagogy, including students analyzing their own genomic data. Motivated by our own medical genomics teaching experiences [27], MySeq enables students to get started performing hands-on genome analyses with just "one click". MySeq can query WGS-scale Tabix-indexed VCF files, either stored locally on the user's computer or remotely available via HTTP(S), without needing to load the entire file. Similar to

GENOtation and DNA Compass, all analyses are performed within the browser without sending any genotypes to a remote server to protect the privacy of users' genomic data. MySeq implements a variety of analyses including variant querying and annotation, physical trait prediction, pharmacogenomics (PGx), polygenic disease risk and ancestry visualization to provide representative pedagogical examples. We describe the implementation of MySeq and our experience employing MySeq in an intensive undergraduate human genome analysis course.

Implementation

MySeq is a single-page web application implemented in JavaScript ES6 with React.js. Figure 1 shows an overview of the dataflow within MySeq. All analyses begin with a compressed and Tabix-indexed VCF file [38]. The user selects a local VCF and its accompanying index file, enters a HTTP(S) URL for a VCF file or selects a preconfigured public genome (NA12878 Genome in a Bottle callset [39]). Alternately the URL of the VCF file can be provided as a URL query parameter. MySeq loads the entire Tabix index (typically 1 MB or less in size) into the browser's memory and uses that index to efficiently determine and load just the small portion of the VCF file containing the variants needed for an analysis. The index calculations, fetch, decompression and VCF parsing are performed entirely within the browser.

MySeq supports the GRCh37/hg19 and hg38 reference genomes and VCF files with multiple samples. The analyses, and particularly the variant annotation functionality, assumes the VCF file is normalized to make all variants bi-allelic, left-aligned and trimmed [40]. A

Fig. 1 Overview of dataflow in MySeq. The MySeq single-page web application performs personal genome analyses in the user's web browser. (1) MySeq components query a locally stored or remotely available VCF file by genomic coordinates. (2) Internally MySeq uses the Tabix index to fetch and parse only the portion of the file containing variants in the query region. (3) MySeq further analyzes the VCF records entirely in the browser (e.g. displays the genotypes to the user, performs ancestry analysis, etc.). Optionally MySeq can utilize the publicly available MyVariant. info and MyGene.info APIs [37] to annotate variants or translate gene symbols or rsIDs to genomic coordinates for queries (e.g. query for all variants in *BRCA1*), but does not send any genotypes to a remote server

normalization script is included in the source repository to assist in preparing data for use with MySeq.

Table 1 describes the functionality currently available in MySeq. Each analysis is implemented as a separate React component. Figure 2 shows the user interface for the VCF loading, variant query and Warfarin PGx components as examples. An analysis component typically queries for one or more variants by genomic position when it loads, dynamically updating the user interface (UI) as the data is returned. The queries are performed in a separate web worker to not block the UI. Since many analyses use similar methods, e.g. mapping the genotypes for a variant to the corresponding phenotypes, a set of shared analysis components are provided for common operations. New analyses can be readily composed from these building blocks.

MySeq does not require its own application-specific server; any HTTP(S) server that supports serving file ranges can be used with MySeq (e.g. Apache or a service like Amazon AWS). MySeq uses the publicly available MyVariant.info API [37] to annotate variants with the predicted amino acid translation, population frequency, links to public databases like ClinVar and other data, and the MyVariant.info and MyGene.info APIs to translate dbSNP rsIDs and gene symbols to genomic coordinates for queries. Only site-level data, e.g. variant position and alleles, and not genotypes (i.e. the alleles present in a specific sample) are sent to a remote server to maintain the privacy of the user's genomic data. The user can optionally block the use of third-party APIs.

Table 1 Description of current MySeq functionality

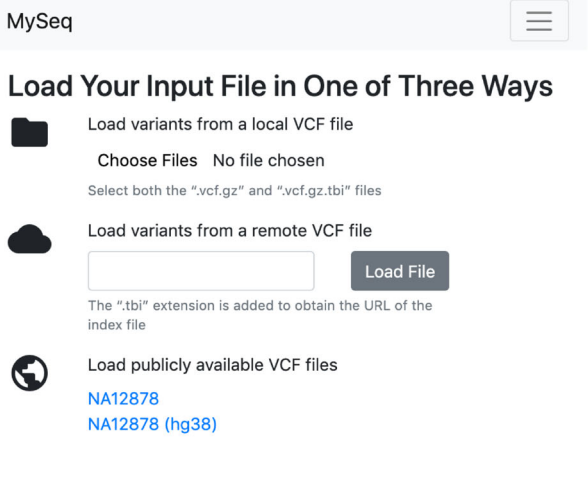| Category	Current Capabilities
Query	Query for variants by genomic position, rsID and gene name. Obtain comprehensive annotations, e.g. amino acid translation, allele frequency, etc., for each variant from MyVariant.info [37].
Physical Traits	Predict phenotypes for physical traits, e.g. taste PTC as bitter, based on the genotypes of one or more associated variants.
Pharmacogenomics (PGx)	Report genotypes and associated phenotypes based on CPIC guidelines and/or the drug label for Simvastatin and Warfarin.
Polygenic Disease Risk	Predict polygenic disease risk for Type 2 Diabetes using a likelihood ratio-based approach [41], and report the *APOE* ε4 genotype and associated risk data for Alzheimer Disease.
Ancestry	Scatter plot visualization of the top two principal components using loadings for 496 ancestry informative markers (AIM) [42] computed with the POPRES resource [43].

The user selects among the available analyses using "client-side routing" so that each analysis component has a unique URL (switching between analyses within the application does not require reloading the VCF file index). By providing a URL to a remote VCF file as a query parameter to an analysis URL, instructors (and others) can distribute links to a specific analysis of specific data.
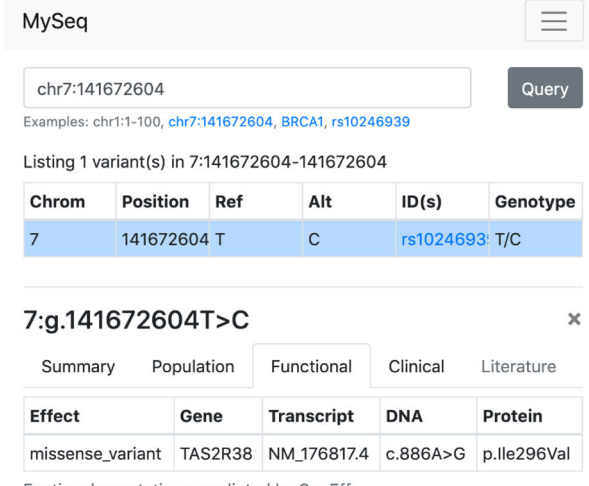
Results

The complexity of genome informatics, and particularly the extensive use of command-line software tools, creates barriers to the wider adoption of experiential genomics education. Creating sustainable genomics pedagogy that can be used in many different educational settings, including those with fewer resources, will require minimizing the need for specialized software and other computational infrastructure [44]. Motivated by the needs we observed in our own genomics teaching we developed MySeq to: 1) enable hands-on personal genome analysis using only the learner's web browser; 2) ensure users can maintain complete control over their genomic data by storing it locally on their computer; and 3) support diverse pedagogy, including independent exploration, structured laboratory exercises and interactive demos.

We employed MySeq in an intensive undergraduate human genome analysis course. Students analyzed both anonymous reference data (the Illumina Platinum Genomes NA12878 trio [45]) and identified personal genome sequencing data individuals had made publicly available via OpenHumans.org [46]. The VCF files were made available via HTTPS on an institutional file server enabling students to get started just by clicking on a link to MySeq that automatically loaded the relevant genome. No file downloads, software installation or other preparatory steps were required.

Students made extensive use of the query functionality to perform their own analyses as part of an independent final project. Example uses included finding and annotating possible disease-causing variants (e.g. in known disease genes) and retrieving the genotype for variants previously reported in the literature. Students completed instructor-created laboratory exercises, e.g. predicting ABO blood group or comparing polygenic disease risk for parents and children, using the relevant scientific literature and links to specific variant queries or other MySeq analyses. These links, or even the MySeq application itself, can be embedded into another webpage to create online demos. An example "demo" that embeds MySeq (via an iframe) and IGV.js [35] to predict whether NA12878 tastes the chemical PTC as bitter (a popular in-class experiment) is available at https://go.middlebury.edu/myseq-demo. Several similar demos using MySeq were integrated into the course materials
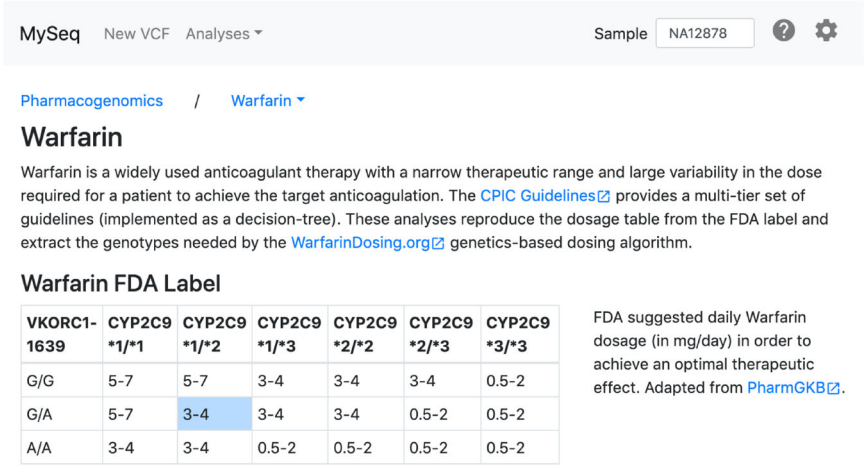
Fig. 2 Example of MySeq VCF loading, variant query and PGx interfaces. **a** The user can load data is several ways, including pre-configured publicly available genomes. **b** Having loaded NA12878's genome, the user's query of chr7:141672604 returned one overlapping variant 7:g.141672604 T > C for which NA12878 is heterozygous. The user clicked on the variant to obtain functional and other annotations from MyVariant.info [37]. (c) Via the "Analyses" dropdown in the header bar (shown fully expanded in the larger screenshot), the user can launch other analyses, e.g. extract variants associated with Warfarin dosing

as interactive complements to the lecture slides and other course materials.

MySeq reduced the computational barriers to learning in this course. The instructor could distribute links to pre-configured analyses of specific data for laboratory exercises and demos that students could use immediately without needing to install or learn to use additional software packages. Instead of just being static demonstrations, these interactive exercises were the starting point for students' independent analyses (again with no additional software required).

The browser-based approach introduces limitations: the scale of the analyses are restricted to an amount of data that can be reasonably downloaded and an amount of computation that be performed within the browser, and most existing genome analysis software would be need to be ported (and likely extensively modified) to work in the browser environment. However, as MySeq and other browser-based tools show, sophisticated analyses are possible, even within those limitations. The flexibility and ease-of-use of "client side" web applications make this an attractive approach for expanding access to experiential genomics education.

Linderman *et al. BMC Medical Genomics*   (2019) 12:172

Page 5 of 7

By supporting both locally stored and remotely available VCF files from within a browser-based tool, MySeq can take advantage of the ease-of-use of a web application while ensuring users can maintain control of their data by only storing it locally. Simply storing data locally, however, does not guarantee security and privacy. MySeq does not provide additional encryption beyond that employed by the user and thus is not a substitute for implementing data security best practices, such as local data encryption.

Conclusion

The growing deployment of genome sequencing in research, clinical and commercial contexts is creating a corresponding need for a more genomically literate workforce and populace. To meet that need we must improve genomics education at all levels. We define "student" broadly. Patient/participant genomic literacy is equally important to the effective application of genomic testing [47]. With many patients/participants now able to obtain their own genomic testing data for further self-directed analysis [48–51], we see a critical need to offer hands-on genomic education to the general public. The most useful pedagogical approaches will be those that can be readily adapted to other educational settings, including those outside traditional academic medical centers, with fewer specialist, infrastructure, and financial resources.

MySeq is not intended however to diagnose, prevent or treat any disease or condition (including to predict a person's response to specific medications). That warning is displayed within the application when loading a VCF file and in the documentation. At present the regulatory "picture" for "third party" tools is unclear and evolving (see [52] for a recent review). Similar to GENOtation [53], the purpose of MySeq is not to perform third-party interpretation, instead MySeq is intended as a hands-on pedagogical tool for learning about how genome analyses are performed.

Here we described MySeq, a single page web-application for personal genome analysis designed to support experiential genomics education. By replacing command-line and other specialized personal genome analysis software with an easy-to-deploy and easy-to-use web application, MySeq makes hands-on personal genome analysis more accessible for students of all kinds. We hope that such a tool will contribute to the larger effort improve the availability and efficacy of genomics education for providers and patient/participants alike.

Availability and requirements

Project name: MySeq.

Project home page: https://github.com/mlinderm/myseq, https://go.middlebury.edu/myseq

Operating system(s): Platform independent.
Programming language: JavaScript.
Other requirements: None.
License: Apache 2.

References
1. Hurle B, Citrin T, Jenkins JF, Kaphingst KA, Lamb N, Roseman JE, et al. What does it mean to be genomically literate?: National Human Genome Research Institute Meeting Report. Genet. Med;15:658–663. https://doi.org/10.1038/gim.2013.14 [cited 2015 Jan 15]
2. Murray MF. Educating physicians in the era of genomic medicine. Genome Med. 2014;6:45 Available from: http://www.ncbi.nlm.nih.gov/pubmed/25031626. [cited 2016 Nov 17].
3. Patay BA, Topol EJ. The unmet need of education in genomic medicine. Am J Med. 2012;125:5–6 Available from: http://www.ncbi.nlm.nih.gov/pubmed/22195527. [cited 2013 Dec 24].
4. Callier SL, Toma I, McCaffrey T, Harralson AF, O'Brien TJ. Engaging the next generation of healthcare professionals in genomics: planning for the future. Per Med. 2014;11:89–98 Available from: http://www.futuremedicine.com/doi/10.2217/pme.13.99. [cited 2016 Nov 22].
5. Plunkett-Rondeau J, Hyland K, Dasgupta S. Training future physicians in the era of genomic medicine: trends in undergraduate medical genetics education. Genet. Med. 2015;17:927–34 Available from: http://www.nature.com/doifinder/10.1038/gim.2014.208. [cited 2016 Nov 22].
6. Bowdin S, Gilbert A, Bedoukian E, Carew C, Adam MP, Belmont J, et al. Recommendations for the integration of genomics into clinical practice. Genet Med. 2016; Available from: http://eresources.library.mssm.edu:2155/gim/journal/vaop/ncurrent/full/gim201617a.html. [cited 2016 May 17].
7. Hooker GW, Ormond KE, Sweet K, Biesecker BB. Teaching genomic counseling: preparing the genetic counseling workforce for the genomic era. J Genet Couns. 2014;23:445–51 Available from: http://www.ncbi.nlm.nih.gov/pubmed/24504939. [cited 2014 Aug 1].
8. Passamani E. Educational challenges in implementing genomic medicine. Clin. Pharmacol. 2013;94:192–5. https://doi.org/10.1038/clpt.2013.38 [cited 2014 Feb 24].
9. Salari K. The dawning era of personalized medicine exposes a gap in medical education. PLoS Med. 2009;6:e1000138 Available from: https://www.ncbi.nlm.nih.gov/pubmed/19707267. [cited 2015 Jan 16].
10. Bennett RL, Waggoner D, Blitzer MG. Medical genetics and genomics education: how do we define success? Where do we focus our resources?

Linderman *et al. BMC Medical Genomics* (2019) 12:172

Page 6 of 7

Genet Med. 2017;19:751–3 Available from: http://www.nature.com/doifinder/10.1038/gim.2017.77. [cited 2019 Jan 2].

11. Garber KB, Hyland KM, Dasgupta S. Participatory genomic testing as an educational experience. Trends Genet. 2016; Available from: http://www.sciencedirect.com/science/article/pii/S0168952516300038.

12. Gerhard GS, Paynton B, Popoff SN. Integrating Cadaver Exome Sequencing Into a First-Year Medical Student Curriculum. JAMA. 2016;315:555–6 Available from: http://jama.jamanetwork.com/article.aspx?articleid=2481601. [cited 2016 Apr 25].

13. Perry CG, Maloney KA, Beitelshees AL, Jeng LJB, Ambulos NP, Shuldiner AR, et al. Educational innovations in clinical pharmacogenomics. Clin Pharmacol Ther. 2016:582–4 Available from: http://doi.wiley.com/10.1002/cpt.352. [cited 2016 Nov 22].

14. Reed EK, Johansen Taber KA, Ingram Nissen T, Schott S, Dowling LO, O'Leary JC, et al. What works in genomics education: outcomes of an evidenced-based instructional model for community-based physicians. Genet. Med. 2016;18:737–45 Available from: http://www.nature.com/articles/gim2015144. [cited 2018 Jun 8].

15. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74 Available from: http://www.ncbi.nlm.nih.gov/pubmed/26432245. [cited 2017 Jul 26].

16. Hyland K, Garber K, Dasgupta S. From helices to health: undergraduate medical education in genetics and genomics. Per. Med. 2018:pme-2018–0081 Available from: http://www.ncbi.nlm.nih.gov/pubmed/30489214. [cited 2018 Dec 25].

17. Talwar D, Tseng T-S, Foster M, Xu L, Chen L-S. Genetics/genomics education for nongenetic health professionals: a systematic literature review. Genet Med. 2017;19:725–32 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27763635. [cited 2019 Jan 2].

18. Talwar D, Chen W-J, Yeh Y-L, Foster M, Al-Shagrawi S, Chen L-S. Characteristics and evaluation outcomes of genomics curricula for health professional students: a systematic literature review. Genet Med. 2018;1 Available from: http://www.nature.com/articles/s41436-018-0386-9. [cited 2018 Dec 25].

19. Rubanovich CK, Cheung C, Mandel J, Bloss CS. Physician preparedness for big genomic data: a review of genomic medicine education initiatives in the United States. Hum Mol Genet. 2018;27:R250–8 Available from: http://www.ncbi.nlm.nih.gov/pubmed/29750248. [cited 2018 Aug 14];.

20. Chrystoja CC, Diamandis EP. Whole genome sequencing as a diagnostic test: challenges and opportunities. Clin Chem. 2014;60:724–33 Available from: http://www.clinchem.org/content/60/5/724.full. [cited 2014 Dec 3].

21. Machluf Y, Gelbart H, Ben-Dor S, Yarden A. Making authentic science accessible-the benefits and challenges of integrating bioinformatics into a high-school science curriculum. Brief Bioinform. 2017;18:145–59 Available from: http://www.ncbi.nlm.nih.gov/pubmed/26801769. [cited 2017 Jan 16].

22. Cummings MP, Temple GG. Broader incorporation of bioinformatics in education: opportunities and challenges. Brief Bioinform. 2010;11:537–43 Available from: http://www.ncbi.nlm.nih.gov/pubmed/20798182. [cited 2017 Jan 16].

23. Adams SM, Anderson KB, Coons JC, Smith RB, Meyer SM, Parker LS, et al. Advancing pharmacogenomics education in the core pharmd curriculum through student personal genomic testing. Am J Pharm Educ. 2016;80:3 Available from: http://www.ncbi.nlm.nih.gov/pubmed/26941429. [cited 2016 Nov 22].

24. Weitzel KW, McDonough CW, Elsey AR, Burkley B, Cavallari LH, Johnson JA. Effects of using personal genotype data on student learning and attitudes in a pharmacogenomics course. Am J Pharm Educ. 2016;80:122 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27756930. [cited 2016 Nov 18].

25. Weber KS, Jensen JL, Johnson SM. Anticipation of Personal Genomics Data Enhances Interest and Learning Environment in Genomics and Molecular Biology Undergraduate Courses. PLoS One. 2015;10:e0133486 Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0133486. [cited 2015 Nov 1].

26. Vernez SL, Salari K, Ormond KE, Lee SS-J. Personal genome testing in medical education: student experiences with genotyping in the classroom. Genome Med. 2013;5:24 Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3706781&tool=pmcentrez&rendertype=abstract. [cited 2013 Dec 24].

27. Linderman MD, Bashir A, Diaz GA, Kasarskis A, Sanderson SC, Zinberg RE, et al. Preparing the next generation of genomicists: a laboratory-style course in medical genomics. BMC Med. Genomics. 2015;8:47 Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4534145&tool=pmcentrez&rendertype=abstract. [cited 2015 Aug 17].

28. Karczewski KJ, Tirrell RP, Cordero P, Tatonetti NP, Dudley JT, Salari K, et al. Interpretome: a freely available, modular, and secure personal genome interpretation engine. Pac Symp Biocomput. 2012:339–50 Available from: http://www.ncbi.nlm.nih.gov/pubmed/22174289. [cited 2016 Nov 23].

29. Curnin C, Gordon A, Erlich Y. DNA Compass: a secure, client-side site for navigating personal genetic information. Bioinformatics. 2017;33:2191–3 Available from: http://www.ncbi.nlm.nih.gov/pubmed/28334237. [cited 2017 Jul 26].

30. Yuan J, Gordon A, Speyer D, Aufrichtig R, Zielinski D, Pickrell J, et al. DNA. Land is a framework to collect genomes and phenomes in the era of abundant genetic information. Nat Genet. 2018;50:160–5 Available from: http://www.nature.com/articles/s41588-017-0021-8 [cited 2019 Jul 1].

31. Miller CA, Qiao Y, DiSera T, D'Astous B, Marth GT. bam.iobio: a web-based, real-time, sequence alignment file inspector. Nat Methods. 2014;11:1189 Available from: http://www.ncbi.nlm.nih.gov/pubmed/25423016. [cited 2016 Dec 15].

32. Ward A, Karren MA, Di Sera T, Miller C, Velinder M, Qiao Y, et al. Rapid clinical diagnostic variant investigation of genomic patient sequencing data with iobio web tools. J Clin Transl Sci. 2017;1:381–6 Available from: http://www.ncbi.nlm.nih.gov/pubmed/29707261. [cited 2019 may 30].

33. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12:996–1006 Available from: https://genome.cshlp.org/content/12/6/996.abstract. [cited 2019 Oct 17].

34. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, et al. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 2016;17:–66 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27072794. [cited 2019 Oct 24].

35. igv.js. Available from: https://github.com/igvteam/igv.js/. Accessed 23 Oct 2019.

36. Vanderkam D, Aksoy BA, Hodes I, Perrone J, Hammerbacher J. pileup.js: a JavaScript library for interactive and in-browser visualization of genomic data. Bioinformatics. 2016;32:2378–9 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27153605. [cited 2019 Oct 24].

37. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, et al. High-performance web services for querying gene and variant annotation. Genome Biol. 2016;17:91 Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0953-9. [cited 2017 Jan 1].

38. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics. 2011;27:718–9 Available from: http://www.ncbi.nlm.nih.gov/pubmed/21208982. [cited 2016 Dec 15].

39. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, et al. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol. 2014;32:246–251. do: https://doi.org/10.1038/nbt.2835. [cited 2014 Feb 19]

40. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. Bioinformatics. 2015;31:2202–4 Available from: http://www.ncbi.nlm.nih.gov/pubmed/25701572. [cited 2019 Jun 3].

41. Morgan AA, Chen R, Butte AJ. Likelihood ratios for genome medicine. Genome Med. 2010;2:30 Available from: http://genomemedicine.com/content/2/5/30. [cited 2012 Mar 10].

42. Drineas P, Lewis J, Paschou P. Inferring Geographic Coordinates of Origin for Europeans Using Small Panels of Ancestry Informative Markers. Relethford J, editor. PLoS One. 2010;5:e11892 Available from: http://www.ncbi.nlm.nih.gov/pubmed/20805874. [cited 2019 Jun 9].

43. Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, et al. The Population Reference Sample, POPRES: A Resource for Population, Disease, and Pharmacological Genetics Research. Am J Hum Genet. 2008;83:347–58 Available from: http://www.ncbi.nlm.nih.gov/pubmed/18760391. [cited 2019 Jun 3].

44. Brazas MD, Lewitter F, Schneider MV, van Gelder CWG, Palagi PM. A quick guide to genomics and Bioinformatics training for clinical and public audiences. PLoS Comput Biol. 2014;10. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003510.

45. Eberle MA, Fritzilas E, Krusche P, Källberg M, Moore BL, Bekritsky MA, et al. A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. Genome Res. 2017;27:157–64 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27903644. [cited 2019 Jan 19].

46. Tzovaras BG, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, et al. Open Humans: A platform for participant-centered research and personal data exploration. bioRxiv. 2019:469189 Available from: https://www.biorxiv.org/content/10.1101/469189v3. [cited 2019 Jun 23].

47. Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011;470:204–13. https://doi.org/10.1038/nature09764 [cited 2014 Jul 9].

48. Ball MP, Bobe JR, Chou MF, Clegg T, Estep PW, Lunshof JE, et al. Harvard Personal Genome Project: lessons from participatory public research. Genome Med. 2014;6:10 Available from: http://genomemedicine.com/content/6/2/10. [cited 2015 Jul 31].

49. Jarvik GP, Amendola LM, Berg JS, Brothers K, Clayton EW, Chung W, et al. Return of Genomic Results to Research Participants: The Floor, the Ceiling, and the Choices In Between. Am J Hum Genet. 2014;94:818–26 Available from: http://www.sciencedirect.com/science/article/pii/S0002929714001815. [cited 2014 May 27].

50. Linderman MD, Nielsen DE, Green RC. Personal Genome sequencing in ostensibly healthy individuals and the PeopleSeq consortium. J Pers Med. 2016;6:14 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27023617.

51. National Academies of Sciences E and M. Returning Individual Research Results to Participants: Guidance for a New Research Paradigm. Downey AS, Busta ER, Mancher M, Botkin JR, editors. National Academies Press; 2018. Available from: https://books.google.com/books?id=AB5sDwAAQBAJ. Accessed 23 Oct 2019.

52. Guerrini CJ, Wagner JK, Nelson SC, Javitt GH, AL MG. Who's on third? Regulation of third-party genetic interpretation services. Genet Med. 2019: 1–8 Available from: http://www.nature.com/articles/s41436-019-0627-6. [cited 2019 Oct 24].

53. Nelson SC, Fullerton SM. "Bridge to the Literature"? Third-Party Genetic Interpretation Tools and the Views of Tool Developers. J Genet Couns. 2018; 27:770–81 Available from: http://www.ncbi.nlm.nih.gov/pubmed/29411211. [cited 2019 Jan 24].

Publisher's Note