The Brazilian Population Tumor Atlas Project: addressing representation in cancer genomics in Brazil

Mariana Boroni,[a,*] Liliane Costa Conteville,[a] Luis Felipe Ribeiro Pinto,[b] and João P. B. Viola[c]

[a]Laboratory of Bioinformatics and Computational Biology, Division of Basic and Experimental Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil
[b]Program of Molecular Carcinogenesis, Division of Basic and Experimental Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil
[c]Program of Immunology and Tumor Biology, Division of Basic and Experimental Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil

The Lancet Regional Health - Americas 2026;56: 101397

Published Online xxx https://doi.org/10.1016/j.lana.2026.101397

Summary

The underrepresentation of Brazil and other Latin American nations in global genomic research is a consequence of multifaceted challenges. Persistent underfunding and infrastructure limitations have hampered the development and implementation of large-scale genomic studies in medicine. This imbalance creates a significant barrier to realizing the potential of precision medicine for the Brazilian population. The disparity in genomic data is particularly concerning in Brazil, given its genetically diverse population with unique ethnic characteristics. This critical need for representative data is precisely why the Brazilian Population Tumor Atlas (*Atlas Tumoral da População Brasileira*, ATPBR) was conceived. Funded by public health initiatives, ATPBR aims to comprehensively map the molecular signatures of cancer and promote responsible data sharing, supporting a future of more equitable care. Its primary objective is to establish a secure platform for the storage and dissemination of multi-omics tumor data. ATPBR has the potential to strengthen Brazil's role in advancing cancer genomics and contributing to local and global cancer control.

Copyright © 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Cancer genomics; Precision oncology; Tumor atlas; Multi-omics

Twenty-two years ago, the Human Genome Project (1990–2003)[1,2] delivered the first complete human reference genome, establishing a foundation for studying genetic variation and disease. This project anticipated an era of medicine where diseases would be decoded, treatments personalized, and all patients served. However, more than two decades later, this vision remains unfulfilled, particularly in low- and middle-income countries.

Building on the foundation of the Human Genome Project, large-scale initiatives such as The Cancer Genome Atlas (TCGA),[3] the International Cancer Genome Consortium (ICGC),[4] and UK Biobank[5] have propelled cancer research through extensive sequencing and data sharing. By systematically sequencing and analyzing tumor samples, both from primary and metastatic sites, and in many cases, matched normal samples across diverse cancer types, these and other projects have built a comprehensive multi-omics resource. By integrating genomics, transcriptomics, epigenomics, and proteomics, these resources provide researchers with deep insights into the molecular mechanisms driving cancer biology.

Over the years, the knowledge generated has paved the way for groundbreaking studies on novel drugs, biomarkers, and innovative strategies for diagnosing, preventing, and treating cancer. Additionally, it has enabled comparisons across different patient cohorts and facilitated large-scale pan-cancer studies, which examine multiple cancer types simultaneously to uncover shared molecular mechanisms, regardless of tissue or organ origin.

Prior to these initiatives, cancer classification relied predominantly on histopathology and limited molecular markers. Now, we have fundamentally shifted this paradigm by defining tumors based on comprehensive molecular landscapes, revealing new insights into tumor heterogeneity and evolution. These insights have been pivotal in advancing precision oncology, facilitating biomarker-driven therapies, and the development of novel drug strategies. The identification of critical mutations, such as IDH in gliomas,[6] and the characterization of molecular subtypes, such as in melanoma,[7] enabled the development of tailored treatments, improving patient outcomes. These projects have also been crucial in advancing immunotherapy by providing

*Corresponding author. Laboratório de Bioinformática e Biologia Computacional, Instituto Nacional de Câncer (INCA), Rua André Cavalcanti, 37, Centro, Rio de Janeiro, RJ, 20231-050, Brazil.
E-mail address: mariana.boroni@inca.gov.br (M. Boroni).

Disclaimer: A translated version of this summary is available in Portuguese in the Supplementary Materials.

insights into tumor mutational burden (TMB), neoantigens, and immune microenvironment interactions, which are now recognized as key determinants in predicting responses to immune checkpoint inhibitors.[8,9] Nonetheless, the vast amount of data and many data types produced by these projects have spurred tremendous growth in the computational biology field and data science.

Notably, Brazil and many Latin American countries have not been at the forefront of this genomic revolution in oncology. While our hospitals address numerous cancer cases daily, the genetic information that guides precision oncology is overwhelmingly derived from populations in the Global North. This reliance on data from other populations can be problematic, as genetic information may not accurately reflect the diversity of our population.[10] Furthermore, evidence indicates that ancestry-related differences in methylation and mRNA expression are causally linked to cancer and cancer-related immune pathways.[11] This vast imbalance underscores the urgent need to develop large-scale genomic studies that specifically include Latin American populations. The generation of region-specific genomic data would allow researchers to improve the accuracy of risk assessments, enhance treatment efficacy, and develop targeted therapies that better address the unique genetic and environmental factors influencing cancer in Latin America.

However, this goal is hindered by the continued underrepresentation of genetic data from Latin American nations in global cancer genomic research, combined with a lack of robust regional initiatives. In Brazil, over the past decades, federal funding for genomic research has consistently prioritised the agriculture sector (Fig. 1A), a strategic sector for national development that generated substantial economic and scientific returns. Although the COVID-19 pandemic prompted an increase in science and technology investments, including the launch of the Genomas Brasil initiative in 2020,[12] this shift has not been sufficient to close the investment gap. An analysis of CNPq funding patterns, one of Brazil's main research funding agencies, reveals that Agronomy, the leading area within Agricultural Sciences, has received significantly more support than Genetics, an important area within Biological Sciences (Fig. 1B). Brazil has demonstrated remarkable capacity in genomic studies, as exemplified by the successful sequencing of the *Xylella fastidiosa* genome, a landmark project[13] that brought considerable benefits to science, the economy, and society.[14] However, this success has not been mirrored in the oncology and precision medicine fields.

Persistent underfunding and infrastructure limitations have not only hampered the development and implementation of local large-scale genomic studies in healthcare, restricting the generation of crucial genomic data relevant to Brazilians, but also hindered the training of specialized personnel in clinical genomics. Key areas such as computational oncology remain underdeveloped, limiting the national capacity to analyze large-scale cancer datasets, build predictive models of tumor progression, and implement Artificial Intelligence (AI)-driven strategies for personalized treatment. Compounding these challenges is the slow progress in establishing national biobanks, which are essential resources for storing and analyzing biological samples. Their development is often obstructed by a range of ethical, regulatory, and logistical hurdles, including informed consent procedures, data privacy concerns, sample transport limitations, and the lack of guidelines that respect the cultural sensitivities of diverse communities within the region.

Furthermore, the paucity of Brazilians and other Latin American populations in global genomic databases creates a substantial knowledge gap, limiting the applicability of findings from other populations and potentially resulting in suboptimal treatment strategies. For example, a study analyzing genetic ancestry across 33 cancer types in a cohort of 9.818 patients from TCGA found that, among the 8.010 participants with available race and/or ethnicity information, 79.01% identified as White (non-Hispanic or unknown ethnicity), while only 4.09% identified as Hispanic/Latino.[15] Notably, these proportions vary by tumor type, but overall reflect a striking lack of representation. This gap is especially concerning given the growing body of evidence showing that population-specific genetic variants can influence cancer risk, prognosis, and therapeutic response.

The disparity in genomic data is particularly concerning in Brazil, given its highly genetically diverse population with unique ethnic characteristics. This diversity, resulting from contributions from European, African, and Indigenous ancestries, along with extensive miscegenation and immigration, has created a genetic "melting pot"[16,17] and offers a unique opportunity to understand cancer genetic features in a diverse population. A recent large-scale genomic study analyzing 2.723 high-coverage whole-genome sequences from diverse Brazilian cohorts identified over 8.7 million previously unreported genetic variants,[18] underscoring the vast, unexplored genomic diversity of Brazilians. Moreover, recent evidence indicates that some of these variants are not adequately captured by existing international or national genetic testing guidelines.[19–21] As a result, current criteria fail to identify between 17% and 25% of carriers of clinically relevant mutations,[19] highlighting the urgent need to tailor genetic screening protocols to better reflect the genetic architecture of underrepresented populations. In addition, some germline variants previously associated with hereditary cancer syndromes show markedly higher allele frequencies in Brazilians compared to global reference datasets (Table 1). A well-known

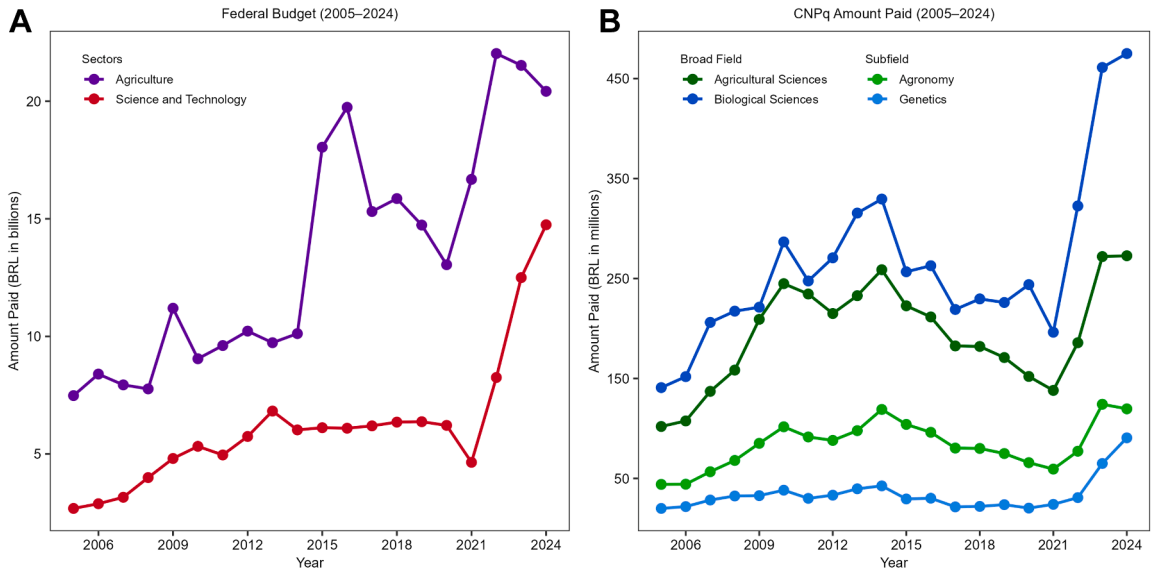

Fig. 1: **Disparities in Brazilian federal investments in Agricultural vs. Biomedical Research in Brazil.** (A) Annual federal budget allocated to the Agriculture and Science and Technology sectors from 2005 to 2024. Data obtained from the Federal Budget Panel available at the SIOP platform–https://www.siop.planejamento.gov.br. (B) Annual funding disbursed by CNPq to projects in Agricultural Sciences and its main subfield Agronomy, as well as Biological Sciences and its main subfield Genetics, from 2005 to 2024. Data retrieved from the CNPq Science, Technology, and Innovation Funding Panel–http://bi.cnpq.br/painel/fomento-cti/index.html.

example is the pathogenic germline variant R337H (c.1010G > A, p.Arg337His) in the *TP53* gene, linked to Li-Fraumeni Syndrome (LFS), a hereditary cancer predisposition syndrome characterized by a high risk of early-onset malignancies. While LFS is considered rare worldwide, the R337H variant exhibits a disproportionately high prevalence in Southern Brazil due to a founder effect.[22] This has led to the implementation of mass newborn screening programs in the state of Paraná.[23]

These findings also expose the limitations of relying on self-reported race and ethnicity as a proxy for genetic ancestry. Studies within Brazil have shown that self-reported skin color/race is often an unreliable

	rs121912664	rs80356898	rs28909982	rs36053993	rs34612342
Gene	*TP53*	*BRCA1*	*CHEK2*	*MUTYH*	*MUTYH*
Variant ID	17-7670699-C-T	17-43093844-G-A	22-28725338-T-C	1-45331556-C-T	1-45332803-T-C
Protein Change	p.Arg337His	p.Gln563Ter	p.Arg117Gly	p.Gly368Asp	p.Tyr151Cys
ClinVar Significance	Pathogenic/Likely pathogenic	Pathogenic	Pathogenic/Likely pathogenic	Pathogenic/Likely pathogenic	Pathogenic/Likely pathogenic
Condition(s)	Li-Fraumeni Syndrome	Breast-Ovarian Cancer	Breast, Prostate, Colorectal Cancer	MUTYH-associated polyposis (MAP)	Colorectal Cancer
Reference	19, 20	20	19, 20	19, 20, 21	21
Brazil AF[a]	0.000367242	0.000183621	0.00128535	0.00642674	0.000918105
gnomAD Genomes AF	0.00001315	0.000006576	0.00008542	0.003349	0.001563
gnomAD Exomes AF	0.000002052	0.000009581	0.0001929	0.003714	0.001959
1000 Genomes AF	–	–	–	0.002396	0.0002013
Fold Enrichment[b] Brazil vs. gnomAD Genomes/Exomes/1000G	27.93/179.06/–	27.91/19.17/–	15.05/6.66/–	1.92/1.73/2.68	0.59/0.47/4.56

[a]Brazilian allele frequencies (Brazil AF) were obtained from the DNA do Brasil project,[18] using data available at: https://lookerstudio.google.com/u/0/reporting/4b3cb1c9-5f26-4278-98d5-9fcb68432ba4/page/p_jsjqk0xugd?s=hRdwUhMcftc. [b]Fold Enrichment was calculated by dividing the allele frequency observed in the Brazilian population (Brazil AF) by the corresponding allele frequency in each international database (gnomAD Genomes, gnomAD Exomes, and 1000 Genomes), calculated independently for each database.

Table 1: **Comparative frequencies of germline pathogenic variants linked to hereditary conditions in Brazil versus global populations.**

predictor of genetic ancestry,[24] highlighting the critical need for robust genetic ancestry markers in pharmacogenomic research and clinical practice. This approach is consistent with the recognition that race and ethnicity are sociocultural constructs rather than biological determinants.

The consequence of this complex genetic landscape is that the prevalence of specific cancer-driving mutations, as well as individual responses to therapies, may differ substantially from those observed in primarily European populations. This disparity is particularly concerning, as less than 2% of NCI-funded clinical trials have included non-White participants, limiting the generalizability of findings and potentially leading to suboptimal treatment outcomes for underrepresented populations.[25]

This critical need for representative data is precisely why ATPBR is more than a mere database. Funded by the Brazilian Ministry of Health in partnership with Brazilian National Council for Scientific and Technological Development (CNPq) in December 2023, ATPBR is a long-awaited scientific and public health initiative aimed at comprehensively mapping the genetic signatures of cancer within our diverse population, ensuring a future of more equitable care. Its primary objective is to establish a secure platform for the storage and sharing of multi-omics tumor data. By centralizing this information on a user-friendly web platform (www.atpbr.com.br), ATPBR will bolster collaboration and data accessibility for researchers nationwide, facilitating a deeper understanding of the molecular underpinnings of various tumors in Brazilians. This repository will house the multi-omics data generated during the project's lifespan, alongside data from parallel initiatives like Genomas Brasil in Cancer (an initiative of the Ministry of Health), which is generating omics data from Brazilian individuals with breast, penile, prostate, oropharyngeal, cervical, ovarian, and endometrial cancers, among others. These malignancies represent a significant public health challenge in the country. A core focus of ATPBR will be the inclusion of underrepresented populations, and representatives from all regions. The inherent genetic diversity of the Brazilian people can influence both the predisposition to and molecular characteristics of these tumors, making in-depth, context-specific investigations essential. In ATPBR, self-reported race/ethnicity will be collected as part of the sociodemographic metadata to support contextual and comparative analyses. However, acknowledging the limitations of this measure, the initiative will also incorporate genomic strategies, whenever possible, to infer ancestry with greater precision. The integration of multi-omics, clinicopathological, and sociodemographic data from patients, combined with machine learning approaches, will enable the identification of clinically relevant diagnostic, prognostic, and treatment response biomarkers. This, in turn, will pave the way for precision medicine within the Brazilian Public Health System (SUS).

To ensure the highest standards of data quality, reproducibility, and comparability, all data submitted to the ATPBR platform will adhere to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) and undergo rigorous quality control procedures. Sample and raw sequencing file inclusion will be contingent upon three mandatory criteria: (i) the presence of high-quality sequencing data, (ii) proper informed consent from the donor, and (iii) the submission of a complete set of minimal metadata fields. These metadata fields include key clinical variables such as cancer classification, disease stage, treatment regimens, and response to therapy, as well as sociodemographic information, including age, sex, self-reported race/ethnicity, and geographic origin. All downstream data processing steps will be conducted using rigorously validated and community-endorsed protocols, including GATK Best Practices for variant discovery,[26] Best Practices for RNA-Seq,[27,28] and established guidelines for epigenomic data processing.[27] To ensure full traceability, version control, and reproducibility, all bioinformatics pipelines will be implemented using workflow management systems based on Nextflow,[29] facilitating data provenance and allowing for consistent reanalysis as methods evolve. Processed multi-omics data, along with clinical annotations, will be made available through interactive dashboards and visual analytics tools, enabling users to explore anonymized, non-sensitive data. Raw sequencing files and intermediate analysis outputs will remain under controlled access and will be made available upon request, subject to review and approval by an independent Data Access Committee.

Beyond national efforts, ATPBR has the potential to enhance collaborations with international consortia, such as the Human Tumor Atlas Network (HTAN),[30] a National Cancer Institute (NCI)-funded Cancer Moonshot SM initiative aimed at constructing three-dimensional atlases of human cancers. Additionally, ATPBR can engage with the Human Cell Atlas (HCA),[31] a global, open-science initiative focused on mapping all human cell types in healthy tissues. Notably, the HCA consortium has made efforts to diversify sample representation worldwide and is now entering its second phase, which aims to develop comprehensive disease atlases, including cancer.

In this context, ATPBR represents a pivotal advancement, positioning Brazil as a key player in the rapidly evolving field of genomics-driven cancer research. However, its long-term success and sustainability depend on a sustained and coordinated commitment to public and private investment. Beyond scientific goals, continued funding must be recognized as a public health priority, essential to ensuring that the benefits of precision medicine are accessible within the

SUS. Immediate priorities include scaling up tumor sequencing efforts, particularly for cancers that disproportionately affect the Brazilian population, while ensuring representation across all five macroregions of the country to capture its genetic diversity and address regional disparities in cancer incidence and outcomes.

Rather than duplicating existing infrastructures, ATPBR is designed to promote integration, interoperability, and data harmonization among public and private biobanks, research centers, and related initiatives. At the same time, it is important to acknowledge that ATPBR addresses only part of a broader challenge: genetic diversity across Latin America reflects distinct Indigenous ancestries, heterogeneous African diaspora contributions, and country-specific colonization and admixture histories, meaning that Brazilian data, despite its diversity, cannot be considered representative of the region as a whole. Accordingly, ATPBR is positioned not as a regional substitute, but as an enabling platform, providing standardized, FAIR-compliant frameworks for data generation, governance, and interoperability that can support future federated analyses and catalyze coordinated, locally driven cancer genomics initiatives across Latin America. By enhancing the use and sharing of existing high-quality data and promoting the generation of new, demographically and regionally diverse datasets, ATPBR seeks to strengthen an inclusive foundation for precision oncology in Brazil, while contributing meaningfully to global cancer genomics and control efforts.

Contributors

MB and JPBV conceived the idea, while MB, JPBV, LFRP, and LCC contributed to the writing, reviewing, and editing of the report. All authors have seen and approved the submitted version.

Statement on the use of AI or AI-assisted technologies

During the preparation of this work the author(s) used ChatGPT in order to improve text readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of interests

MB, JPBV, LFRP are members of the ATPBR project, which is funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Acknowledgements

Funding: MB, JPBV and LFRP received funds from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and LCC from Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lana.2026.101397.

References

1 Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
2 Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304–1351.
3 Hutter C, Zenklusen JC. The cancer genome atlas: creating lasting value beyond its data. *Cell*. 2018;173:283–285.
4 Zhang J, Bajari R, Andric D, et al. The international cancer genome consortium data portal. *Nat Biotechnol*. 2019;37:367–369.
5 Conroy MC, Lacey B, Bešević J, et al. UK Biobank: a globally important resource for cancer research. *Br J Cancer*. 2023;128:519–527.
6 Killock D. CNS cancer: molecular classification of glioma. *Nat Rev Clin Oncol*. 2015;12:502.
7 Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell*. 2015;161:1681–1696.
8 Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18:248–262.
9 Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity*. 2018;48:812–830.e14.
10 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51:584–591.
11 Carrot-Zhang J, Chambwe N, Damrauer JS, et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. *Cancer Cell*. 2020;37:639–654.e6.
12 Brasil, Ministério da Saúde, Departamento de Ciência e Tecnologia (DECIT). Programa Nacional de Genômica e Saúde de Precisão – Genomas Brasil. Brasília: Ministério da Saúde [cited 2025 Jun 22]. Available from: https://www.gov.br/saude/pt-br/composicao/sectics/decit/genomas-brasil.
13 Simpson AJG, Reinach FC, Arruda P, et al. The genome sequence of the plant pathogen Xylella fastidiosa. *Nature*. 2000;406:151–157.
14 Giovanetti M, Slavov SN, Fonseca V, et al. Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. *Nat Microbiol*. 2022;7:1490–1500.
15 Yuan J, Hu Z, Mahal BA, et al. Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer Cell*. 2018;34:549–560.e9.
16 Souza AMD, Resende SS, Sousa TND, Brito CFAD. A systematic scoping review of the genetic ancestry of the Brazilian population. *Genet Mol Biol*. 2019;42:495–508.
17 Kehdy FSG, Gouveia MH, Machado M, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci U S A*. 2015;112:8696–8701.
18 Nunes K, Araújo Castro ESM, Rodrigues MR, et al. Admixture's impact on Brazilian population evolution and health. *Science*. 2025;388:eadl3564.
19 de Oliveira JM, Zurro NB, Coelho AVC, et al. The genetics of hereditary cancer risk syndromes in Brazil: a comprehensive analysis of 1682 patients. *Eur J Hum Genet*. 2022;30:818–823.
20 Sandoval RL, Leite ACR, Barbalho DM, et al. Germline molecular data in hereditary breast cancer in Brazil: lessons from a large single-center analysis. *PLoS One*. 2021;16:e0247363.
21 Ma X, Zhang B, Zheng W. Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Gut*. 2014;63:326–336.
22 Galante PAF, Guardia GDA, Pisani J, et al. Personalized screening strategies for R337H carriers: a retrospective cohort study of tumor spectrum in Li-Fraumeni syndrome adult carriers. *Lancet Reg Health Am*. 2025;42:100982.
23 Achatz MIW, Hainaut P, Ashton-Prolla P. Highly prevalent TP53 mutation predisposing to many cancers in the Brazilian population: a case for newborn screening? *Lancet Oncol*. 2009;10:920–925.
24 Pereira JL, de Souza CA, Neyra JEM, et al. Genetic ancestry and self-reported 'Skin Color/Race' in the urban admixed population of São Paulo City, Brazil. *Genes*. 2024;15. https://doi.org/10.3390/genes15070917.
25 Chen MS Jr, Lara PN, Dang JHT, Paterniti DA, Kelly K. Twenty years post-NIH Revitalization Act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual: renewing the case for enhancing minority participation in cancer clinical trials. *Cancer*. 2014;120(Suppl 7):1091–1096.
26 Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1–11.10.33.
27 Luo Y, Hitz BC, Gabdank I, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res*. 2020;48:D882–D889.

28 Deshpande D, Chhugani K, Chang Y, et al. RNA-seq data science: from raw data to effective interpretation. *Front Genet*. 2023;14: 997383.
29 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 2017;35:316–319.
30 Rozenblatt-Rosen O, Regev A, Oberdoerffer P, et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. *Cell*. 2020;181:236–249.
31 Rood JE, Maartens A, Hupalowska A, Teichmann SA, Regev A. Impact of the human cell atlas on medicine. *Nat Med*. 2022;28:2486–2496.