Perspective

The human and non-human primate developmental GTEx projects

https://doi.org/10.1038/s41586-024-08244-9

Received: 14 June 2024

Accepted: 17 October 2024

Published online: 15 January 2025

Check for updates

Tim H. H. Coorens[1]✉, Amy Guillaumet-Adkins[1], Rothem Kovner[2], Rebecca L. Linn[3], Victoria H. J. Roberts[4], Amrita Sule[1], Patrick M. Van Hoose[5] & the dGTEx Consortium*

Many human diseases are the result of early developmental defects. As most paediatric diseases and disorders are rare, children are critically underrepresented in research. Functional genomics studies primarily rely on adult tissues and lack critical cell states in specific developmental windows. In parallel, little is known about the conservation of developmental programmes across non-human primate (NHP) species, with implications for human evolution. Here we introduce the developmental Genotype-Tissue Expression (dGTEx) projects, which span humans and NHPs and aim to integrate gene expression, regulation and genetics data across development and species. The dGTEx cohort will consist of 74 tissue sites across 120 human donors from birth to adulthood, and developmentally matched NHP age groups, with additional prenatal and adult animals, with 126 rhesus macaques (*Macaca mulatta*) and 72 common marmosets (*Callithrix jacchus*). The data will comprise whole-genome sequencing, extensive bulk, single-cell and spatial gene expression profiles, and chromatin accessibility data across tissues and development. Through community engagement and donor diversity, the human dGTEx study seeks to address disparities in genomic research. Thus, dGTEx will provide a reference human and NHP dataset and tissue bank, enabling research into developmental changes in expression and gene regulation, childhood disorders and the effect of genetic variation on development.

The human body consists of trillions of cells working in harmony, which possess essentially the same genome, yet have distinct morphologies, functionalities and expression patterns. Over the past decade, many efforts have begun to characterize and catalogue adult tissue and cell expression patterns. Starting in 2010, the Genotype-Tissue Expression (GTEx) project[1–3] has established a large resource of gene expression and molecular profiles across a wide variety of human tissues. The large number of donors included in GTEx has enabled researchers to identify inherited genetic variants that alter gene expression through regulatory effects[4–6]. Technological advances since the mid-2010s have enabled the application of single-cell transcriptomic profiling to population-scale cohorts. Rather than aggregating the expression signal derived from bulk tissue, single-cell RNA sequencing enables characterization of cell types and states, including dynamic expression states and differentiation trajectories. Using these technologies, efforts such as the Human Cell Atlas[7], the Human BioMolecular Atlas Program (HuBMAP)[8], PsychEncode[9] and the Brain Initiative Cell Atlas Network (BICAN) are systematically charting human cell types and their transcriptomic profiles across tissues and organs. More recently, spatial sequencing methods have enabled characterization of gene expression in situ[10].

Most of these studies have focused on adult tissues and have not mapped the developmental trajectory in early life across cell types. Starting with the unicellular zygote, cascades of division and differentiation generate the diversity of organs, tissues and cell types that are present in adults[11]. Over the course of fetal, paediatric and adolescent development, cellular phenotypes and expression patterns change drastically before settling into their adult versions. Some recent efforts have been established to generate reference maps of developing tissues but they have limitations. Studies in the Human Developmental Cell Atlas (HDCA)[12,13] have begun to chart expression patterns in human development, focusing on prenatal tissues. However, these studies generally focus on sampling a single tissue type from many donors, rather than multiple tissues from the same donor, which hampers cross-tissue comparisons and investigations into the effect of genetic variation on expression. Moreover, these studies have not systematically and densely sampled different ages from birth to adulthood and may miss critical stages in paediatric and adolescent development.

A rigorous understanding of normal human development is essential to understand the origins of many diseases. Inherited disorders, ranging from congenital malformations[14] and heart defects[15,16] to neurological disorders[17], can have their origins during gestation. Some childhood cancer types are thought to arise from transient, developmental cell types that are ill-represented in adult tissues[18,19]. Furthermore, environmental or genetic influences on early development are associated with diseases that manifest in adult life stages. For instance, epidemiological evidence suggests that children who are malnourished

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Yale University, New Haven, CT, USA. [3]Children's Hospital of Philadelphia, Philadelphia, PA, USA. [4]Division of Reproductive and Developmental Sciences, Oregon National Primate Research Center, Oregon Health and Sciences University, Portland, OR, USA. [5]National Disease Research Interchange, Philadelphia, PA, USA. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: tcoorens@broadinstitute.org

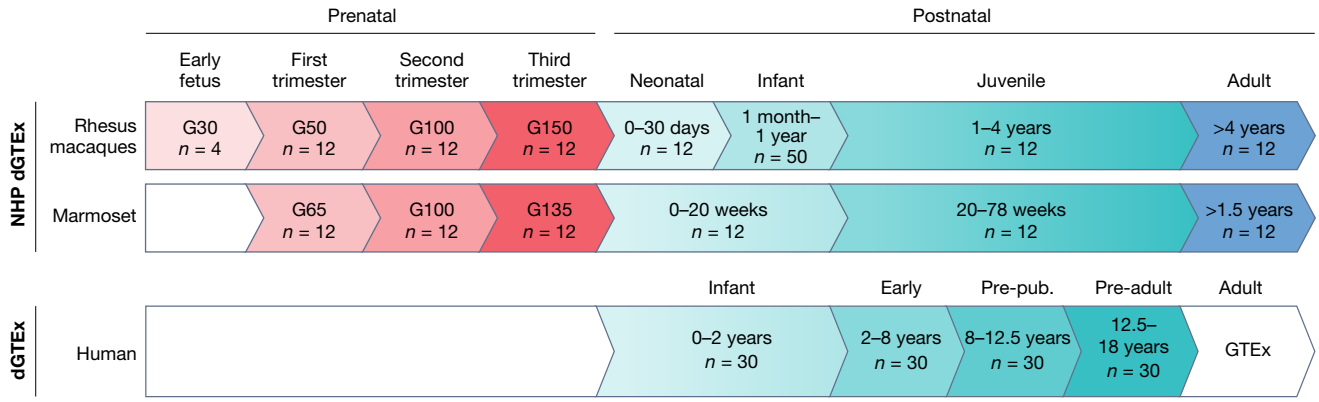

Fig. 1 | Sampling across development and species. Schematic representation of the dGTEx sampling timepoints across the developmental timeline of rhesus macaques (top), marmosets (middle) and humans (bottom). *n* refers to the number of donors per group. The typical age in gestational days (meaning days post-conception) at birth is G165 for macaques, G145 for marmosets and G265 for humans.

during fetal development are at risk for metabolic disorders as adults[20–22]. In addition, the timing of puberty, a trait with a substantial genetic component, is a risk factor for many complex diseases[23,24] later in life.

In parallel, little is known about the precise evolutionary conservation of regulatory networks and developmental programmes. Most mammalian biology has been extrapolated from studies of a few model organisms, some of which are evolutionarily distant from humans and have critical developmental differences. In rodents, a considerable amount of fetal organ development occurs after birth, which limits the usefulness of rodents as a model for the later stages of human pregnancy. Similarly, the developmental trajectory of the early postnatal period differs greatly between rodents and humans. Consequently, our ability to explain many aspects of human development is limited, with failures of some human clinical trials highlighting the inherent problems of translating promising findings in model organisms into human therapies[25]. By contrast, the course of development in NHPs closely mirrors human development[26].

Here we describe the dGTEx project, which aims to expand upon the adult GTEx dataset[2] by profiling tissue-specific gene expression patterns across the course of development and primate species. These projects seek to establish a comprehensive molecular resource database spanning diverse tissues across developmental stages in humans and two NHP species, rhesus macaques and marmosets. In addition, these projects will build unique tissue repositories for future investigations. Through this robust reference of transcription and regulation across early ages and species, dGTEx will enable research into developmental and childhood diseases and disorders, the effect of genetic variation during development, and translational therapies.

Design of the dGTEx projects

The dGTEx projects will sample NHPs and humans across key developmental stages. The human project aims to recover tissues from 120 post-mortem donors, with approximately 30 donors allocated to each of four distinct age groups: infancy, early childhood, pre-pubertal and post-pubertal (Fig. 1). This donor demographic is challenging owing to lower death rates, restrictions of medical examiners or coroners, and authorization for donation. Therefore, sample collection relies on five Organ Procurement Organizations (OPOs) across the USA, which together provide a significant volume of paediatric donor referrals to screen for eligible donors. Donor eligibility relies on satisfying several criteria (Box 1) and uses the existing organ and tissue transplant donor screening processes of the OPOs to provide donor families the opportunity to consider donation for the dGTEx project.

To maximize the scientific and clinical impact of the dataset, a large amount of donor metadata will be collected during donation and biospecimen collection, building on practices developed for GTEx[2]. De-identified donor-level data will include demographic information, medical history, sample-based laboratory test results and death circumstances. Sample-level data will include tissue type, tissue location, ischaemic time and pathology review tissue metrics. A dGTEx-specific questionnaire will capture paediatric-focused metadata from next of kin, including medical history of childhood diseases, birth circumstances, environmental exposure and puberty staging. Additionally, a one-question framework for race and ethnicity reporting is used, as recommended by the Federal Interagency Technical Working Group on Race and Ethnicity Standards convened by the Office of Management and Budget in 2022 (88 FR 5375). This aligns with the Ethical, Legal and Social Implications (ELSI) Diversity, Equity and Inclusion (DEI) goals and will support accuracy of reporting race and ethnicity for dGTEx donors.

The NHP component of dGTEx includes a representative New World species, the common marmoset, and an Old World species, the rhesus macaque. Unusually among NHP species, marmosets have developed unique reproductive systems to reduce the time to sexual maturity, and predominantly produce dizygotic twins that share a placenta[27]. This results in exchange of haematopoietic stem cells between littermates and a lifelong blood chimerism. These characteristics and their relatively small size make the marmoset an increasingly used model organism for biomedical research. Rhesus macaques are the most extensively used NHPs for biomedical research[28]. They are phylogenetically more closely related to humans than marmosets and their reproductive biology is highly similar in terms of uterine and placental structure and physiology. We will sample 126 rhesus macaques and 72 marmoset individuals across developmental stages, spanning both prenatal and postnatal development. All postnatal sampling timepoints are designed to match developmental windows across all three species. The study of NHPs in parallel with the human postnatal timepoints allows for greater understanding of postnatal development across several primate species. In addition, the use of two NHP species enables the analysis of fetal NHP samples, which are challenging to obtain in humans.

Macaque samples will be collected from the animal colony at the Oregon National Primate Research Center (ONPRC), and marmoset samples will be obtained from the colony at Massachusetts Institute of Technology (MIT). Fetal tissue will be collected at known gestational ages as part of the time-mated breeding programmes at ONPRC and MIT. Eligible postnatal animals that are otherwise underutilized for behavioural, social or medical reasons, are identified for necropsy and tissue collection (Box 2). Within the developmental cohorts, animal numbers will be balanced by sex. Cohort sizes are selected to be conservative in our use of precious animal resources.

Box 1

Inclusion and exclusion criteria for human donors

dGTEx inclusion criteria for all donors

1. Donor is 18 years of age or under.
2. Tissue collection can begin within 24 hours of cross-clamp and/or cardiac cessation (observed or presumed).
3. Donor did not receive whole-blood transfusion within the previous 48 hours.
4. Donor does not have current positive blood cultures (sepsis).
5. Donor has never been diagnosed with metastatic cancer (cancer that spread beyond the initial site, such as to other organs like the brain or liver).
6. Donor has not received chemotherapy or radiation therapy for cancer in the past two years.
7. Donor does not have a history of intravenous drug abuse within the past five years.
8. Donor does not have a history of HIV/AIDS, hepatitis C virus or hepatitis B virus or a history of exposure to HIV/AIDS, hepatitis C virus or hepatitis B virus through needlesticks, contact with non-intact skin, contact with open wounds and/or contact with mucous membranes; if the donor is less than 18 months of age or has been breastfed in the past 12 months, the birth mother must meet this condition.
9. Donor was neither hospitalized because of COVID-19 nor died of COVID-19.
10. Donor has no known chromosomal disorder (for example, Down syndrome).
11. Donor has no history of failure to thrive or total parenteral nutrition.

dGTEx exclusion criteria for brain donation

1. Donor cause of death related to penetrating brain injury or head trauma.
2. Donor was ventilator-dependent for more than 24 hours.

Box 2

Inclusion and exclusion criteria for NHPs

NHP dGTEx inclusion criteria for all NHPs (macaques and marmosets)

1. NHP is within two days of age bracket if prenatal.
2. Tissues can be collected immediately after euthanasia.
3. The majority of tissues are available for collection.
4. NHP has no clinical or laboratory evidence of sepsis or other systemic infection.
5. NHP has no known genetic or chromosomal disorder or evidence of congenital malformations.
6. NHP is not genetically modified and has not received gene therapy treatment.

dGTEx exclusion criteria for brain donation

1. NHP has a history of head trauma or neurologic signs.

Ethical, legal and social implications

The American Society for Human Genetics has noted that "addressing underrepresentation in human genomics starts with meaningful engagement of underrepresented communities"[29]. dGTEx seeks to address this with an integrated ELSI-focused substudy, in which the study team will engage with the broad scientific groups throughout all phases of the project. This integrated approach will assess the unique ELSI aspects of dGTEx by implementing a model called DEI 360°, which includes engaging geographically, racially and ethnically, and socio-culturally diverse stakeholders from the inception of the project. Current stakeholders consist of family decision-makers, tissue requesters, community advisory board members, paediatric healthcare professionals and a geographically diverse team of paediatricians and paediatric health psychologists throughout all phases of dGTEx. Feedback from community stakeholders has already informed key study aspects, including how race and ethnicity will be assessed and reported, content of study documents (for example, donor authorization form), supplemental family resources (for example, family-facing website), and development of educational materials for tissue requesters. This approach will continue to inform the study at multiple levels and will offer continued opportunities to revise processes. The integrated ELSI substudy will seek to outline gold standards for community engagement, thereby enhancing donor diversity, ensuring culturally appropriate donation requests and shaping best practices for future genomic research.

Tissue sampling

The tissue sampling in the dGTEx projects uses a common coordinate framework[30–32] to enable reproducible sampling across sites, donors and species, where anatomically feasible. The collected tissue sites and metadata are aligned with other consortia and datasets such as the HDCA, HuBMAP, BICAN and adult GTEx to facilitate the integration of dGTEx results with those efforts.

Sampling for the human dGTEx component will be performed by recovery teams as soon as possible following recovery for transplant. To provide consistency across all sites, recovery teams undergo specific tissue recovery training on detailed standard operating procedures for procurement, packaging and shipping of the tissue samples. Instruction diagrams and images have been developed as visual guides for the recovery teams during tissue procurement. On-site instruction by a paediatric pathologist and National Disease Research Interchange team members is available for OPOs throughout the project. The tissue collection and sampling schema will provide biospecimens for molecular assays within dGTEx and a biobank to enable future studies with novel technologies, thus expanding the impact of each tissue donation for both donor families and the scientific community.

The tissue sampling schema varies on the basis of organ size and structure. In general, a small sample (2.5 cm × 1 cm × 1 cm cube for solid organs and 2.5 cm × 1 cm × full thickness for mucosal organs) will be collected from most tissue sites. Each sample will be divided into two frozen aliquots for molecular analysis and one fixed aliquot for histopathologic evaluation (Fig. 2). Organ size changes during development pose several challenges for common coordinate framework tissue sampling. For example, the intestinal length changes significantly during development. In adult GTEx, the jejunum was sampled at a fixed distance from the ligament of Treitz, but the same sample site will be located differently depending on donor age in the dGTEX cohort.

Similarly, some organs (for example, ovary, testis, adrenal gland and thyroid gland) are small in younger donors and will be collected whole to obtain sufficient tissue for downstream molecular analysis. For bilateral organs, recovery of both organs provides enough tissue for molecular analysis, histopathologic evaluation and biobanking. Tissue collection protocols were developed to maximize high-quality

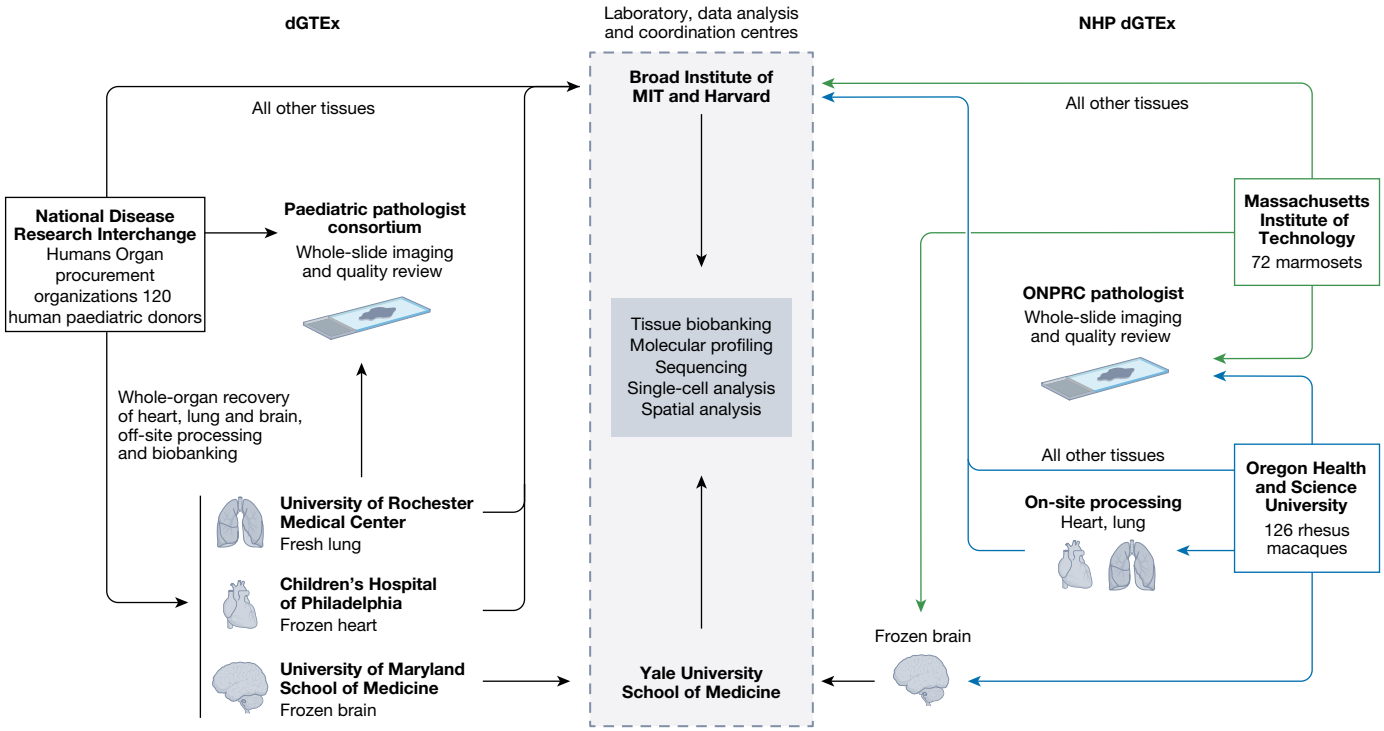

Fig. 2 | Overview of the dGTEx workflow. Schematic detailing the workflow for human (left) and NHP (right) post-mortem donors, tissues and aliquots, showing collection sites, processing locations and methods, and the data analysis and biorepository centres.

tissue recovery despite the above challenges, with an expected 20 to 74 different tissue sites sampled per donor, depending on organ donation, authorization of tissues and restriction of medical examiners and coroners. To maximize tissue availability, we have developed informative resources for OPOs to engage in proactive education efforts with medical examiners and coroners.

When available, the heart, lungs and brain will be recovered whole and further dissected into subregions. If the entire heart is available, bilateral atria will be sampled on site and the remainder of the heart will be frozen. A paediatric pathologist from the Children's Hospital of Philadelphia will conduct further heart sub-dissections. The left and right lung will be collected fresh and sent to the University of Rochester Medical Center to be sub-dissected following established LungMAP tissue protocols[33–35]. Brain will be dissected into cerebellum, brainstem and two hemispheres. One hemisphere will be further sub-dissected for downstream molecular analyses following established protocols used in BICAN. For consistency, all brain sub-dissections for the human and NHP species will be conducted at Yale School of Medicine.

Marmoset and macaque tissue sampling is designed to map to the human tissue sampling as closely as possible, with modifications for species-specific anatomical and developmental differences. For example, NHPs have an additional right lung lobe and a different hepatic lobation pattern, and lack a sigmoid colon. NHP tissue availability will not be affected by organ donation and will include additional tissues that are challenging to obtain in humans (eyes, thymus, cerebrospinal fluid and heart valves) and prenatal tissues (placenta, umbilical cord, fetal membranes and amniotic fluid). Importantly, there will be minimal tissue ischaemia since sample collection at ONPRC or MIT begins immediately following euthanasia.

The tissue size of NHPs can be challenging, especially among marmosets. Sampling schemas have been modified to whole organ or whole-body collections for small animals at the earliest timepoints (Fig. 3), which impose constraints on sub-dissection (Supplementary Table 1). Because NHPs are a limited and valuable biomedical research resource, maximal use of their tissues and organs is imperative. For gestational day 30 (G30) macaques and G65 and G100 marmosets, we will collect whole bodies. In G50 rhesus and G135 marmosets, fewer individual organ dissections are feasible owing to small body size and technical challenges, despite the use of a dissection microscope by trained veterinary personnel. When possible, prenatal samples will be bisected for both histology and molecular purposes. However, if unfeasible because of size, frozen tissue for molecular analysis is prioritized.

Following protocols established for adult GTEx[1,2], subspecialty pathologists will review fixed tissue samples to validate tissue origin, content and quality. Tissue sections will be processed and scanned at the University of Maryland Brain and Tissue Bank (brains from both NHP and human), at Children's Hospital of Philadelphia (human, non-brain tissue) and ONPRC (NHP, non-brain tissue). Pathologists will review and report on multiple parameters, including confirming that the correct target tissue is present, the degree of autolysis, the presence of non-target tissue in sample, and any unexpected findings (for example, neoplasia, inflammation or infection). These reports will be made available to the downstream molecular analysis groups to inform assay design and provide real-time feedback to the tissue procurement teams for process improvement.

Assays and analysis

Whole-genome sequencing will be performed on all human and NHP donors to provide an individual germline sequence for use in analyses, including the assessment of molecular effects of inherited genetic variation[2], to improve genome and isoform structure annotation, and to enable genetic demultiplexing of batched samples. Bulk tissue RNA sequencing will also be performed on most tissues, as a relatively cost-effective means to characterize transcript diversity across tissues, individuals and species, and to enable data integration with existing large datasets.

Taking advantage of many novel technologies, some tissues will also be more deeply characterized at the cellular level using single-nucleus

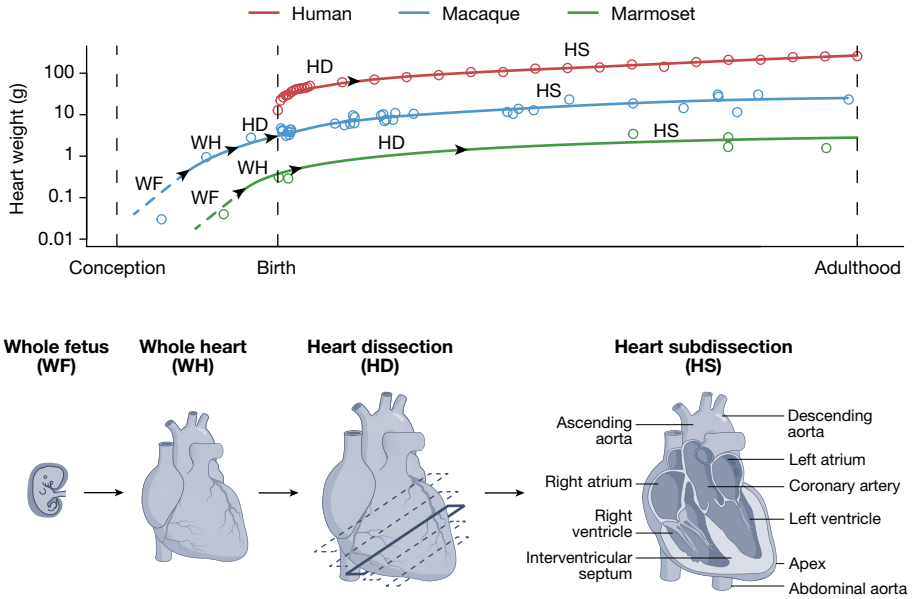

Fig. 3 | Heart sampling strategy across age ranges and species. Heart weights for humans[34,35], macaques[73] and marmosets across age groups sampled in dGTEx. The solid line indicates a locally estimated scatterplot smoothing regression, black solid arrowheads signify the transition to a more detailed sampling strategy, as indicated in the overview below the graph.

RNA sequencing (snRNA-seq), single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq), multiomic RNA/ATAC single-cell analyses, spatial transcriptomics assays, Hi-C and chromatin precipitation with sequencing (ChIP–seq). In addition to short-read assays, we will apply multiplexed arrays isoform sequencing (MAS-ISO-seq)[36] long-read sequencing of transcripts to enhance our understanding of developmental and tissue-specific isoforms[37] and add to the ongoing effort to discover new genes in macaques and marmosets, all of which will enhance our understanding of their genetic and functionally comparative landscapes. The selection of the biospecimens on which these methods are applied will be flexible, depending on the needs of the consortium, resource constraints and any early findings. At minimum, the selection will include tissues across the major organ groups (that is, germ layers) of at least one male and one female across the defined age groups across the three species.

snRNA-seq allows profiling of the transcriptome of individual cells from high-quality primary and frozen tissues[7,38]. Single-cell ATAC-seq (scATAC-seq)[39,40], designed to identify open chromatin regions in the genome, is a key assay to understand regulatory elements, decipher cellular diversity and understand cell decision-making[41–43]. Understanding identities and states of cells is substantially increased by analysing both gene expression and chromatin accessibility within individual cells. This can be achieved experimentally through multiomic techniques that provide both data types from the same cells[44], or computationally by merging information from different experiments or technologies[45]. Joint analysis of chromatin accessibility and gene expression enables better annotation of cell identities[46], the inference of regulatory regions that interact such as enhancers and promoters, and the comparison of expression with accessibility to transcription factors[47].

In addition, it is crucial to understand how cells interact and form spatially structured tissues and microenvironments, which may change during development. The integration of spatial transcriptomic data with pathology provides valuable datasets in histological imaging studies. This will provide insight into the spatial patterns linked to cellular organization, tissue structure and cell–cell communication[48]. Spatial transcriptomics can also offer a solution for very small tissues that cannot be sub-dissected, such as the prenatal NHP samples. Technologies developed for gene expression profiling vary in spatial resolution, efficiency and sensitivity, and are constantly changing, with recent high-resolution, affordable, transcriptomic-based methods.

The dGTEx project aims to generate snRNA-seq, scATA-seq and spatial transcriptomic data from multiple tissues and species across developmental stages. Within our evolving experimental design and a landscape of rapidly changing technologies, we will consider multiple ways to prioritize and stratify samples for analysis. The comprehensive nature of our sampling schema for a broad number of organs collected from a well-sized, sex-balanced cohort provides several analysis design options: (1) a cross-tissue analysis that uses multiple tissues from a few individuals; (2) tissue-specific analyses that use selected tissues from a larger number of individuals; or (3) comparative analyses of sex-, age- and development-specific differences. Irrespective of design, we are committed to generating robust data using protocols and pipelines harmonized with other existing efforts to produce a tool and data resource[49] that can be readily used by the research community.

Analytical challenges and opportunities

A central goal of dGTEx is to capture gene expression patterns and evaluate chromatin accessibility and structure across three diverse species, spanning developmental stages and tissue types. This presents considerable challenges and opportunities. First, reference annotations are incomplete, as some cell types, developmental states or expressed transcripts may not have been previously characterized across all species. Most comparative functional research in NHPs has been constrained to analysing bulk gene expression and regulatory patterns within limited tissues. The absence of comprehensive references for both NHP species may complicate the alignment and annotation of sequencing reads, affecting accurate cell-type identification, gene expression analyses and cross species comparisons. However, recent advances in generating marmoset[50] and macaque[51] genomes using long-read assembly, including expected telomere-to-telomere references[52], show promising robustness compared with previous efforts. With the data that will be generated as part of these projects, especially the long-read RNA sequencing, we can address these gaps in references.

Second, data integration across developmental stages and species[53–55] requires appropriate normalization and batch correction while preventing loss of biological variability[56,57]. Although this may be a challenge[58,59], the uniformity of sample processing within the unique span of dGTEx across development and species also provides a robust dataset to assess existing and develop novel computational tools for the integration and analysis of expression data. A key aspect in addressing these challenges is rapid release of the data to the wider scientific community to test, adapt and further develop analytical methods. Therefore, the dGTEx projects will release the data on a regular basis throughout the timespan of the Consortium.

A strength of the dGTEx projects is that pipelines and references will be fully harmonized with other ongoing efforts by a cross-project working group, including BICAN, and HuBMAP. In this project, characterization of prenatal development and fetal tissues is limited to the NHP species for which scarcity of similar datasets will present additional cell annotation challenges. Fetal tissues, which undergo rapid developmental changes, exhibit diverse cell populations and dynamic gene expression, posing significant analytical challenges in understanding these complex cellular dynamics at a single-cell level. Comparative analyses with fetal human datasets[60,61] will be needed to identify conserved and divergent features during prenatal development.

Whereas dGTEx provides a unique opportunity to discover and characterize genetic regulatory effects in developing tissues, its sample size may limit the statistical power for standalone genetic analysis within dGTEx. Furthermore, analysis of genetic effects that are active only at specific developmental stages further requires modelling of the changing effect size. To boost statistical power and disease applications, genetic analyses in dGTEx will benefit from novel statistical methods (for example, in deep learning), as well integration as with other genetic datasets[62–64].

Outcomes

The RNA-sequencing data generated in the dGTEx projects will enable a robust characterization of transcriptomic profiles across diverse tissues within and across individuals, with the additional axes of development and species adding pivotal resolution and dimensions to the data resource. As well as bulk expression data, the dGTEx projects will rely on single-cell sequencing approaches, both for transcriptomic and epigenomic profiling, especially to identify certain cell types that disappear over development. Although cell-type plasticity in childhood may be less marked than during embryonic or fetal development, the precise profiles and changes of paediatric cells are profoundly important to our understanding of normal development and disease. For instance, many childhood cancers are thought to arise from undifferentiated, transient cells in development[18,19]. Additionally, specific cell types are thought to lose their ability to self-renew and transition into post-mitotic, differentiated cells over the course of development, with implications for studying ageing and senescence.

A unique value of the dGTEx project is the integration of genetic variation with expression profiles. This integration enables us to assess the effects of individual polymorphisms on the patterns of splicing or gene expression, in specific tissues and cell types. Moreover, some effects of inherited genetic variation may be transient in early life and only affect specific human developmental stages, but they may have echoes throughout the remainder of life and can potentially explain established links between genetic variation and phenotypic traits. Genetic insights from dGTEx will not be limited to variant effects detected in dGTEx donors alone, as overlaying existing catalogues of disease-associated variants with molecular annotations from dGTEx will enable inference of disease-relevant cell types and developmental stages.

Given the greater differences seen between species relative to those within a single species, differences that distinguish one species from another tend to have stronger phenotypic effects compared with variations within a species[65]. This makes comparative studies a powerful tool for identifying specific genetic areas that are worth further exploration in humans. Human accelerated regions (HARs) are areas of the genome with significant human-specific sequence changes implicated in evolution. HARs have been identified to be functionally relevant during important prenatal developmental periods in the brain (HAR1 (ref. 66)) as well as in other organs (*HACNS1* (ref. 67)). In essence, comparative genomics and population genetics together within primate functional genomics research offers an exclusive opportunity to pinpoint human-specific variations. This, in turn, can substantially enhance our understanding of various human traits and their evolution.

Another significant project component will be to store all residual collected tissues for future research purposes and further characterization, in the same manner as the biobank created the adult GTEx biospecimens. This resource, to consist of frozen and histological samples, holds great value for subsequent research studies to further enhance our understanding of biological processes in normal tissues spanning a wide age range from well-characterized subjects. This biospecimen collection will be available for other large-scale projects to extend their scope to paediatric and NHP development.

Tissue collection and data generation are currently ongoing. The open access data and pipelines from the dGTEx projects will be made available online on the GTEx Portal (https://gtexportal.org), with periodic updates as data is released. The GTEx Portal will provide the ability to download the public data generated by the dGTEx projects and to view selected data in interactive visualizations across the projects. The rapid release of dGTEx data is likely to fuel research in areas outside the immediate scope of the project, as was the case for adult GTEx, which formed the basis for research into X-inactivation[68,69], cancer expression profiles[70], somatic mutations in normal tissues[71,72] and many other areas.

Conclusion

Paediatric tissues are not simply smaller versions of their adult counterparts but have different cell-type compositions and physiology. Because of this developmental variation in gene expression, children are susceptible to a unique set of diseases and have different, often poorly understood, responses to treatments that were developed for and tested in adults. Moreover, little is known about the evolutionary conservation and differences of developmental programmes between humans and NHPs. To address these gaps, the human and NHP dGTEx project will create a unique database of gene expression and regulation across two temporal axes (development within, and comparatively across, species). Further, the opportunity to have dGTEx target well-known gaps in diversity, equity and inclusion within genomic research with a novel, integrated ELSI-focused approach will create more robust findings and advance their broad applications and impact.

Importantly, dGTEx can provide the groundwork and infrastructure for many future research studies. Such initiatives could include elucidating the origins of paediatric diseases, including childhood cancers and developmental disorders; assessing the consequences of rare congenital syndromes and inherited genomic alterations on the typical course of development; evaluating the effect of gene perturbations and drug treatments on NHP expression patterns; and understanding the evolutionary conservation of development. Together, the dGTEx projects will provide a reference dataset and tissue bank for human and NHP development, enabling research into developmental and childhood disorders, the effect of genetic variation during development, genetic causes of developmental disorders, and translational therapies.

1. Lonsdale, J. et al. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
This paper outlines the original GTEx project, the predecessor of the dGTEx projects.
2. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
3. eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* **49**, 1664–1670 (2017).
4. Kim-Hellmuth, S. et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).
This paper illustrates the variation of genetic effects on gene regulation between cell types from GTEx data.
5. Ferraro, N. M. et al. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science* **369**, eaaz5900 (2020).
6. Oliva, M. et al. The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).
7. Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The Human Cell Atlas: from vision to reality. *Nature* **550**, 451–453 (2017).
This paper outlines the conceptual framework for the Human Cell Atlas, a comprehensive reference map of cell types through application of single-cell technologies to dissect cellular functions, interactions and tissue organization, with significant implications for advancing our understanding of human biology.
8. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
9. Emani, P. S. et al. Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024).
10. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
11. Shahbazi, M. N. Mechanisms of human embryo development: from cell fate to tissue shape and back. *Development* **147**, dev190629 (2020).
12. Haniffa, M. et al. A roadmap for the Human Developmental Cell Atlas. *Nature* **597**, 196–205 (2021).
13. Taylor, D. M. et al. The Pediatric Cell Atlas: defining the growth phase of human development at single-cell resolution. *Dev. Cell* **49**, 10–29 (2019).
14. Mukhopadhyay, N. et al. Whole genome sequencing of orofacial cleft trios from the Gabriella Miller Kids First Pediatric Research Consortium identifies a new locus on chromosome 21. *Hum. Genet.* **139**, 215–226 (2020).
15. Zaidi, S. et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
16. Priest, J. R. et al. De novo and rare variants at multiple loci support the oligogenic origins of atrioventricular septal heart defects. *PLoS Genet.* **12**, e1005963 (2016).
17. Short, P. J. et al. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616 (2018).
18. Coorens, T. H. H. & Behjati, S. Tracing and targeting the origins of childhood cancer. *Annu. Rev. Cancer Biol.* **6**, 35–47 (2022).
19. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat. Med.* **25**, 367–376 (2019).
20. Zheng, X. et al. Risk of metabolic syndrome in adults exposed to the great Chinese famine during the fetal life and early childhood. *Eur. J. Clin. Nutr.* **66**, 231–236 (2012).
21. Ning, F. et al. Famine exposure in early life and risk of metabolic syndrome in adulthood: Comparisons of different metabolic syndrome definitions. *J. Diabetes Res.* **2019**, 7954856 (2019).
22. Grey, K. et al. Severe malnutrition or famine exposure in childhood and cardiometabolic non-communicable disease later in life: a systematic review. *BMJ Glob. Health* **6**, e003161 (2021).
23. Day, F. R. et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017).
24. Day, F. R. et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat. Commun.* **6**, 8842 (2015).
25. The elephant in the room. *Nature* **444**, 790 (2006).
26. Tardif, S., Carville, A., Elmore, D., Williams, L. E. & Rice, K. in *Nonhuman Primates in Biomedical Research* Vol. 1 (eds Abee, C. R. et al.) 197–249 (Elsevier, 2012).
27. Tardif, S. D. & Ross, C. N. in *The Common Marmoset in Captivity and Biomedical Research* (eds Marini, R. et al.) 119–132 (Academic Press, 2019).
28. Lewis, A. D. & Prongay, K. in *The Nonhuman Primate in Nonclinical Drug Development and Safety Assessment* (eds Bluemel, J. et al.) 87–113 (Academic Press, 2015).
29. Lemke, A. A. et al. Addressing underrepresentation in genomics research through community engagement. *Am. J. Hum. Genet.* **109**, 1563–1571 (2022).
This paper highlights the need to address underrepresented populations within genomic research and how community engagement can aid in confronting this issue.
30. Rood, J. E. et al. Toward a common coordinate framework for the human body. *Cell* **179**, 1455–1467 (2019).
31. Burger, A. et al. Towards a clinically-based common coordinate framework for the human gut cell atlas: the gut models. *BMC Med. Inform. Decis. Mak.* **23**, 36 (2023).
32. Börner, K. et al. Construction and usage of a human body Common Coordinate Framework comprising clinical, semantic, and spatial ontologies. Preprint at https://doi.org/10.48550/arXiv.2007.14474 (2020).
33. Ardini-Poleske, M. E. et al. LungMAP: the molecular atlas of lung development program. *Am. J. Physiol.* **313**, L733–L740 (2017).
34. Reid, W. Height and weight in human beings. Autopsy report. *J. Clin. Pathol.* **41**, 237–237 (1988).
35. Schulz, D. M., Giordano, D. A. & Schulz, D. H. Weights of organs of fetuses and infants. *Arch. Pathol.* **74**, 244–250 (1962).
36. Al'Khafaji, A. M. et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat. Biotechnol.* **42**, 582–586 (2024).
37. Glinos, D. A. et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* **608**, 353–359 (2022).
38. Grün, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810 (2015).
39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
40. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
41. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
42. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
43. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
44. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nat. Methods* **17**, 11–14 (2020).
45. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
46. Wang, S. K. et al. Single-cell multiome of the human retina and deep learning nominate causal variants in complex eye diseases. *Cell Genom.* **2**, 100164 (2022).
47. Bravo González-Blas, C. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
48. Cang, Z. et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* **20**, 218–228 (2023).
49. Castel, S. E. et al. A vast resource of allelic expression data spanning human tissues. *Genome Biol.* **21**, 234 (2020).
50. Mao, Y. et al. Structurally divergent and recurrently mutated regions of primate genomes. *Cell* **187**, 1547–1562 (2024).
51. He, Y. et al. Long-read assembly of the Chinese rhesus macaque genome and identification of ape-specific structural variants. *Nat. Commun.* **10**, 4233 (2019).
52. Makova, K. D. et al. The complete sequence and comparative analysis of ape sex chromosomes. *Nature* **630**, 401–411 (2024).
53. Zhang, X. et al. Towards a primate single-cell atlas. *Zool. Res.* **43**, 691–694 (2022).
54. Han, L. et al. Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*. *Nature* **604**, 723–731 (2022).
55. Qu, J. et al. A reference single-cell regulomic and transcriptomic map of cynomolgus monkeys. *Nat. Commun.* **13**, 4069 (2022).
56. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
57. Sikkema, L. et al. An integrated cell atlas of the lung in health and disease. *Nat. Med.* **29**, 1563–1577 (2023).
58. Hrovatin, K. et al. Integrating single-cell RNA-seq datasets with substantial batch effects. Preprint at *bioRxiv* https://doi.org/10.1101/2023.11.03.565463 (2023).
59. Rosen, Y. et al. Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. *Nat. Methods* **21**, 1492–1500 (2024).
60. Cao, J. et al. A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
61. Zhai, J. et al. Primate gastrulation and early organogenesis at single-cell resolution. *Nature* **612**, 732–738 (2022).
62. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
63. van der Wijst, M. et al. The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020).
64. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
65. Housman, G. & Gilad, Y. Prime time for primate functional genomics. *Curr. Opin. Genet. Dev.* **62**, 1–7 (2020).
66. Pollard, K. S. et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
67. Prabhakar, S. et al. Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
68. Tukiainen, T. et al. Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
69. Werner, J. M., Ballouz, S., Hover, J. & Gillis, J. Variability of cross-tissue X-chromosome inactivation characterizes timing of human embryonic lineage specification events. *Dev. Cell* **57**, 1995–2008.e5 (2022).
70. Wang, Q. et al. Unifying cancer and normal RNA sequencing data from different sources. *Sci. Data* **5**, 180061 (2018).
71. Yizhak, K. et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, eaaw0726 (2019).
72. Rockweiler, N. B. et al. The origins and functional effects of postzygotic mutations throughout the human life span. *Science* **380**, eabn7113 (2023).
73. Roberts, V. H. J. et al. Rhesus macaque fetal and placental growth demographics: A resource for laboratory animal researchers. *Am. J. Primatol.* **85**, e23526 (2023).

Acknowledgements This research is supported by the National Human Genome Research Institute (NHGRI), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute of Mental Health (NIMH) and the Office of Research Infrastructure Programs under awards U24 HD106537, U24 HG012090, U24 HG012108 and U24 HG012483. T.H.H.C. is supported by an EMBO long-term fellowship (ALTF 172-2022). We thank the donors and their families for making this study possible. The views and opinions expressed in this manuscript are those of the authors only and do not necessarily represent the views, official policy or position of the US Department of Health and Human Services or any of its affiliated institutions or agencies.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08244-9.
Correspondence and requests for materials should be addressed to Tim H. H. Coorens, Thomas Bell, Kristin G. Ardlie, Nenad Sestan or Donald F. Conrad.
Peer review information *Nature* thanks Rachel Freathy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Biospecimen Procurement Center (BPC) and Organ Procurement Organizations (OPOs): U24HD106537
Thomas Bell[5] ✉, Thomas Blanchard[6], Raquel Hernandez[7], Rebecca Linn[3], Deanne Taylor[3], Melissa VonDran[5], Taha M. Ahooyi[3], Danette Beitra[8], Anas Bernieh[9], Meghan Delaney[10], Melissa Faith[7], Emmanouel Fattahi[11], Dana Footer[10], Michelle Gilbert[12], Simoné Guambaña[7], Sam Gulino[10], Jade Hanson[7], Emilie Hattrell[5], Casie Heinemann[13], Joseph Kreeb[14], Daniel Leino[12], Laurel Mcdevitt[10], Abigail Palmieri[15], Mary Pfeiffer[16], Gloria Pryhuber[14], Chrisopher Rossi[10], Immanuel Rasool[11], Russell Roberts[12], Ahmad Salehi[13], Emmitt A. Savannah[12], Kristen Stachowicz[17], David Stokes[3], Lawrence Suplee[17], Patrick Van Hoose[5], Benjamin J. Wilkins[3], Schawnte' Williams-Taylor[12] & Shiping Zhang[3]

Laboratory, Data Analysis, and Coordinating Center (LDACC)-Broad: U24HG012090
Kristin G. Ardlie[1] ✉, Gad Getz[1,18,19], Tuuli Lappalainen[20,21], Stephen B. Montgomery[22], François Aguet[1], Lisa Anderson[1], Brad Bernstein[1,19,23], Abhishek Choudhary[1], Tim H. H. Coorens[1], Laura Domenech[1], Elizabeth Gaskell[1], Amy Guillaumet-Adkins[1], Matthew Johnson[1], Qiuyue Liu[1], Andrew R. Marderstein[22], Jared Nedzel[1], Joseph Okonda[1], Evin M. Padhi[22], MaryKate Rosano[1], Andrew J. C. Russell[1], Amrita Sule[1] & Brady Walker[1]

Laboratory, Data Analysis, and Coordinating Center (LDACC)-Yale: U24HG012108
Nenad Sestan[2] ✉, Mark Gerstein[2], Aleksandar Milosavljevic[24], Beatrice Borsari[2], Hyesun Cho[2], Declan Clarke[2], Ashley Deveau[2], Timur Galeev[2], Kevin Gobeske[2], Irbaz Hameed[2], Anita Huttner[2], Matthew Jensen[2], Yunzhe Jiang[2], Rothem Kovner[2], Jiaqi Li[2], Jia Liu[2], Yuting Liu[2], Jay Ma[2], Shrikant Mane[2], Ran Meng[2], Anandita Nadkarni[2], Pengyu Ni[2], Saejeong Park[2], Varduhi Petrosyan[24], Sirisha Pochareddy[2], Iva Salamon[2], Yan Xia[2], Chris Yates[24], Menglei Zhang[2] & Hongyu Zhao[2]

Non-Human Primate (NHP)-dGTEX: U24HG012483
Donald F. Conrad[4] ✉, Kristin G. Ardlie[1], Guoping Feng[25], Nenad Sestan[2], Fritzie Brady[4], Magalie Boucher[25], Lucia Carbone[4], Jenna Castro[4], Ricardo del Rosario[1], Madison Held[4], Jon Hennebold[4], Ariah Lacey[25], Anne Lewis[4], Ana Cristina Lima[4], Eisa Mahyari[4], Samantha Moore[25], Mariam Okhovat[4], Victoria Roberts[4], Samia Silva de Castro[25], Brady Wessel[4], Heather Zaniewski[1] & Qiangge Zhang[25]

National Institutes of Health (NIH)
Alexander Arguello[26], Jacob J. Baroch[26], Jyoti Dayal[26], Adam Felsenfeld[26], John V. Ilekis[27], Sheethal Jose[26], Nicole C. Lockhart[26], Daniel Miller[28], Mollie Minear[27], Melissa Parisi[27], Amanda Price[29], Erin Ramos[26] & Sige Zou[30]

[6]University of Maryland School of Medicine, Baltimore, MD, USA. [7]Johns Hopkins All Children's Hospital, Baltimore, MD, USA. [8]Nicklaus Children's Hospital, Miami, FL, USA. [9]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [10]Children's National Hospital, Washington, DC, USA. [11]Infinite Legacy, Halethorpe, MD, USA. [12]Life Gift Organ Donation Center, Fort Worth, TX, USA. [13]Donor Network West, San Ramon, CA, USA. [14]University of Rochester Medical Center, Rochester, NY, USA. [15]Center for Organ Recovery and Education, Pittsburgh, PA, USA. [16]ConnectLife, Williamsville, NY, USA. [17]Gift of Life, Philadelphia, PA, USA. [18]Massachusetts General Hospital, Boston, MA, USA. [19]Harvard Medical School, Boston, MA, USA. [20]Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. [21]New York Genome Center, New York, NY, USA. [22]Stanford University, Stanford, CA, USA. [23]Dana-Farber Cancer Institute, Boston, MA, USA. [24]Baylor College of Medicine, Houston, TX, USA. [25]Massachusetts Institute of Technology, Cambridge, MA, USA. [26]National Human Genome Research Institute, Bethesda, MD, USA. [27]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA. [28]National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA. [29]National Institute of Mental Health, Bethesda, MD, USA. [30]Office of Research Infrastructure Programs, NIH Office of Director, Bethesda, MD, USA. ✉e-mail: tbell@ndriresource.org; kardlie@broadinstitute.org; nenad.sestan@yale.edu; conradon@ohsu.edu