Chapter

Perspective Chapter: From Embryonic Archives to Translational Insights – Reusing Nonhuman Primate Brain Data to Decode Development and Reduce Animal Sacrifice

Alvaro Duque, Phil Barello, Lucy Greene, Emma Burke, Valeria Mendoza-Silva, Yury M. Morozov and Pasko Rakic

Abstract

A serious challenge when conducting studies of neurogenesis and brain development in nonhuman primates (NHPs) is the lack of sufficient, homogeneous data and methodologies to assure accurate, repeatable studies. Matching of age, sex, histological processes, and planes of orientation is paramount, and hence, standards need to be set as a basis for comparisons between normal controls and specimens with lesions, treatments, or pathologies. Part of the challenge is that during development, everything is rapidly changing. Even a few days' mismatch in embryonic age can translate into sizable differences in structural maturity in some brain areas. Sexual dimorphism in the embryonic rhesus macaque is evidenced by differences in average body and brain size and weight at birth. Hence, comparisons of cortical and subcortical regions, glia, neurons, and axonal pathways necessitate the normalization of data with enough matching specimens to ensure meaningful and accurate averages. Natural interindividual variability plays a role and increases the statistical variance in otherwise identical samples. As a solution to this problem, we propose an increase in the sharing and archiving of specimens and data, which ultimately results in an increase in the number of available samples for analysis. This is achievable by building well-documented, accurate, and reliable embryonic datasets. This chapter introduces the creation of such datasets within Collection 6 of the MacBrain Resource Center (MBRC) as a standard for assuring transparency, accuracy, reliability, and repeatability of developmental NHP brain studies. This effort is based on the extended use of materials collected from unrelated studies.

Keywords: rhesus macaque, embryo, MacBrain Resource Center, study reproducibility, decreased animal use, 3Rs

1. Introduction

Most studies of brain development using macaques are rather limited in terms of the number of animals and the completeness of samples per age, sex, and histological repertoire. Repeated imaging of the fetal brain by ultrasound or MRI usually requires repeated anesthesia of the mother, which introduces anesthetic agents, stress, and anxiety as confounding factors. These methods may be globally informative but insufficient. Repeated imaging may, therefore, be impractical and may, in addition, not provide sufficient cytoarchitectonic resolution. Histological processing of one sample is not repeatable in the same specimen at different ages, but it is unsurpassed in terms of cyto-resolution. Obviously, C-sections are required for the collection of the specimens, and if a particular sex is desirable, a priori determination of the sex is also needed. In any case, accurate embryonic age is challenging because it requires precision in the timing of pregnancy and very accurate measurement of crown-rump lengths as a proxy for age if imaging is being used to estimate it. Once an embryo is collected, brain tissue processing is not precisely easy because the samples are usually very fragile, and if fixation is limited to immersion, a proper balance between fixation strength, exposure time, and quality histo- and immunohistochemistry needs to be achieved. The production of embryos requires a breeding colony, and the collection of embryos requires experienced and specialized veterinary medical personnel able to perform C-sections and subsequently care for the recovery of the mother.

Given the difficulties in procuring NHP embryos for studies of brain development, the building of accurate and reliable datasets with reusable information seems not only justifiable but necessary to foster and facilitate developmental brain research. The building of such a database is challenging and expensive, but it is underway within the MacBrain Resource Center (MBRC) in the Department of Neuroscience at the Yale University School of Medicine. Making all data publicly available constitutes true data sharing that encourages data reuse, promotes discovery, enhances scientific rigor, and diminishes costs, both in terms of animal lives and in terms of the financial burden of data production. Some fundamental aspects of scientific importance are the proper documentation of data collection and characterization and the proper and safe storage of the data while still assuring a user-friendly interface. These aspects are crucial for data reuse and repeated analysis using ever-changing analytical tools.

Here, we describe our efforts to provide the scientific community with such a database. As rhesus macaque embryos have been collected for unrelated studies, we get what is left over and not needed by those original studies, and process available brain tissue further in an organized and systematic way. Our data repository is intended to be the already established and operational Collection 6 of the MBRC (https://macbraingallery.yale.edu/collection6/). For completeness, the descriptions below cover from specimen procurement to examples of data analysis.

2. Procurement of the specimens

2.1 The breeding colony: Production of timed pregnancies

The production of timed pregnancies is a process that starts with daily vaginal swabs. In time, the females get accustomed to the process, and with the help of positive reinforcement, they readily present. The collection is done with the tip of a cotton applicator for the detection of menstrual blood, which is subsequently used to estimate

the time of ovulation. The female in heat is paired with the male for ~72 hours, that is, for three nights, starting around the morning of the 10th day after the onset of menstruation. The male and female are separated on the morning of the 4th day. During this time, the pair usually copulates several times. In the case of a proven pregnancy, the approximate time of conception is estimated to be the middle of the second day of pairing, so our uncertainty is ~ ±1.5 days. This can, of course, be longer given the possible long survival of the sperm before the fertilization of the egg. The estimate has, however, over the years, proved to be accurate enough for our purposes.

2.2 The determination of sex

When necessary, the determination of embryo sex is done by testing maternal blood serum for the presence of the DNA sex-determining region of the Y gene (SRY). This is done in-house by polymerase chain reaction (PCR) following the protocol described by Mitsunaga et al., [1]. To increase accuracy, blood from the pregnant macaque is preferably obtained ~35 days or later after pairing.

2.3 Collection of embryos

C-sections are performed by qualified veterinarians at the embryonic age specified by research needs (see **Figure 1**). A team of experts performs the surgeries following the same high medical standards followed in the human procedure. There is always a dedicated surgeon and a dedicated anesthesiologist. Several veterinary medical assistants help monitor vital signs and handle surgical instruments. Pre- and postoperative care is also managed by professionals, which always include veterinary residents who conduct daily clinical rounds. The husbandry staff is highly qualified to recognize and report any signs of distress. After euthanasia by sodium pentobarbital overdose, larger embryos are usually transcardially perfused with 0.1 M phosphate buffer saline (PBS), followed by 4% paraformaldehyde (PF) and 15% picric acid in PBS (Zamboni fixative). In the case of smaller embryos, the whole head or even the whole body may be fixed by immersion in the same fixative. In these cases, the fontanelle is usually opened with the tip of a surgical blade to aid the fixative solution in reaching the brain. Immersion is usually overnight, and the same is done for older brains, but postfixation may be longer depending on the quality of the primary fixation. Cryoprotection is done in 20–30% sucrose in PBS for as long as needed until the brain sinks in the solution. Timing depends on the size of the brain, which is a function of age.

2.4 Brain cutting and tissue storage

After proper cryoprotection, brains are either blocked or stored whole. They are frozen in isopentane surrounded by dry ice and stored at −80°C. To produce different series of sections that can be systematically processed for different histological markers, we prefer to cut sections at 50 microns thick in a cryostat (Leica CM1950) (See **Figure 2**). Most of the tissue is cut in the coronal plane, but there are instances in which it is cut in sagittal or axial orientations. When a complete brain or a complete hemisphere is available, we cut the whole rostral-to-caudal extent, and depending on the size and number of sections that are acquired (which depends on brain size determined by age), we select a set number of series to be processed so that each series is processed for a single marker, and the sections within each series are equidistant. For instance, from a 15 mm brain, we obtain 300 coronal sections. In a case like

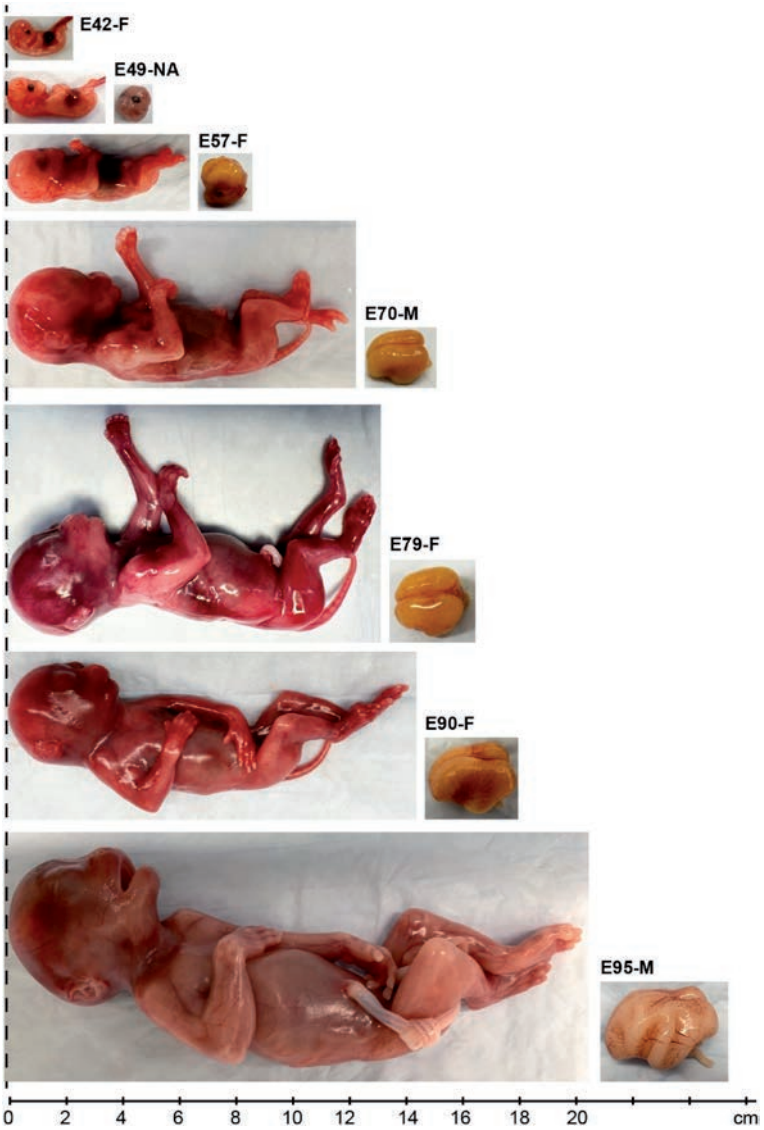

Figure 1.
Examples of rhesus macaque embryos of different ages and sexes. Seven specimens are arranged top to bottom by age ± 1.5 days. Images of the head or brain are also provided from E49 to E95; all images are to scale. The large variability of body sizes is easily appreciated and helps understand the need to collect several specimens of the same sex and age in order to obtain averaged data more truly representative of the population. Abbreviations: male (M), female (F), not available (NA).

this, one can separate 10 series of 30 sections each. For younger embryos with smaller brains, it may be necessary to collect more specimens, since the number of different labels that can be done in one brain is limited by brain size. Sections that are too far apart may fail to properly sample a small developing structure. Too many sections per series decrease the number of different series that can be obtained per specimen. We try to balance the sampling rates to maximize the amount of information that can be obtained per specimen while, at the same time, trying to decrease the number of specimens that are needed. Since the specimens that make it to the MBRC collections

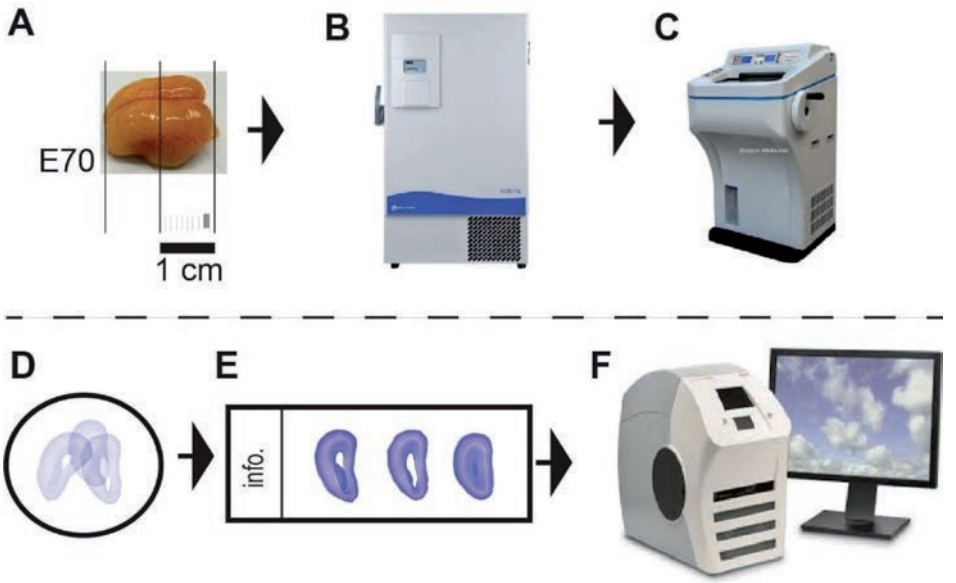

Figure 2.
Storing, cutting, staining, glass mounting, and digitization of brain sections. Fixed and cryoprotected brains (A) are stored at −80°C (B) until cut in a cryostat (C). Sections were then processed free-floating (if possible) (D) and mounted on glass slides (E). After further processing and cover slipping, they are digitized (F).

are left over from unrelated experiments, it may be the case that only one hemisphere is available or that just a few blocks of tissue are available; whatever the case, we try to maximize the amount of tissue that can be processed in an effort to, in time, decrease the need to sacrifice additional animals.

2.5 Histo- and immunohistochemistry processing

Over the years, our histological processes have been described in detail in different publications [2–12]. All histological processing is done free-floating unless it is not possible, such as may be the case with the youngest specimens, in which case they are processed on the slide. All solutions are prepared in PBS. Sections are always rinsed 2x or 3x in PBS and endogenous peroxidases inactivated with 0.2–0.5% H_2O_2 for about 2 and to 20 minutes. After further rinsing, sections are incubated in solutions containing the primary antibody for 24–36 hours at room temperature (RT). Proper dilutions are empirically predetermined and vary from antibody to antibody. Typical dilutions are 1:1000 to 1:10,000. Oftentimes, the solution with the primary antibody also contains 0.2% Triton X to improve antibody penetration. After rinsing, sections are incubated in appropriate biotinylated secondary antibodies. Typical dilutions are 1:200 to 1:500, and the solution typically contains 2–5% serum from the host of the secondary antibody and may also have the same concentration of bovine serum albumin (BSA). All these steps are intended to decrease nonspecific background staining, resulting in a better signal-to-noise ratio. Incubation is usually at RT for 4–8 hours. After rinsing, a final incubation is done in avidin-biotin-peroxidase complex (Vectastain elite ABC kit, Vector labs, Burlingame, CA) for 4 hours at RT, following the manufacturer's recommended dilution. The development of the immunoreaction product is done with 3′,3′-diaminobenzidine as chromogen precipitated by H_2O_2 [10, 12–14]. After

washing, the sections are mounted on microscope glass slides, dried, dehydrated in a series of alcohols, cleared in xylenes, and cover-slipped with Permount® (Fisher Scientific, Fair Lawn, NJ). In general, embryonic tissue is much more delicate than adult tissue. It has less neuropil and a higher water content. Adjustments are usually necessary on a case-by-case basis to preserve tissue integrity and improve quality.

2.6 Imaging and digitization of sections and digital data storage

The MBRC performs ScanScope Virtual Slide (SVS, slide and viewable storage) image file ingest using the Aperio CS2 image capture device, driven by Aperio ScanScope software, to store digitized microscope slides at up to 40x magnification. Physical microscope slides are digitally scanned, and the SVS image files are organized in a database, stored in our network-accessible storage repository, and accessible in an Aperio eSlideManager web interface that includes a full-resolution viewer capable of displaying whole-slide images through a web browser to authenticated visitors. Alternatively, Aperio ImageScope software is an option limited to devices running a Microsoft Windows operating system. A monthly fee is incurred based on the amount of data stored. University information technology support (ITS) experts oversee data integrity and security. If necessary, slides can also be scanned by a different scanner, such as the Zeiss Axioscan 7, an Olympus VS200, or a Huron TissueScope.

2.7 Construction of public web galleries of whole-slide images

Because the Aperio CS2 HR scanner has been our primary instrument in the digitization of data, the descriptions that follow are targeted to this system. However, the principles are the same, independent of the system employed.

2.7.1 Digital input

Aperio ScanScope Virtual Slide (SVS) images are big-tagged image file format (TIFF) files "fully compatible with the TIFF standard, with no proprietary extensions" [15]. Whereas the original TIFF image standard allowed for a maximum file size of 4 GB, BigTIFF images use 64-bit offsets, supporting image sizes up to 18 exabytes (EB).

SVS images include image details, with slide image data stored using lossy JPEG2000 compression and a macro view in low resolution of the slide label stored with lossless Lempel–Ziv–Welch (LZW) compression. The first image in the set of images contained within an SVS file is always a full-resolution tiled baseline image, usually 240 × 240 pixels [15], and the second image is a 1024 × 768-pixel stripped thumbnail of the macro, followed by a set of one or more intermediate pyramidal images, which are compressed with the same algorithm as the base image, and an optional final slide label image.

Leica Biosystems maintains Aperio eSlideManager, which was previously developed by Aperio Technologies, Inc. Aperio eSlideManager is a digital pathology system of record.

2.7.2 Conversion process

Individual SVS image files can be located within the storage repository using the ImageID value displayed within the eSlideManager web interface, originally assigned

by the eSlideManager database. To make Aperio SVS images more accessible, without requiring visitors to install software plugins or provide login credentials, the MBRC leveraged OpenSlide, a software interface to whole-slide images in multiple image formats, including SVS [16]. Deep Zoom image format is XML-based and supports single images, sparse images with multiple regions having different dimensions, and collections of multiple images [17]. Each Deep Zoom image is stored as a tiled pyramid with a 2-dimensional (height and width) base. Every level of the pyramid, from the tip of the peak with the fewest pixels at level 0 to the lowest level, which varies by image height and width values, is a 256-tile image.

Each Aperio SVS image file can be converted to Deep Zoom image format using libvips, an open-source image processing library compatible with OpenSlide image files. OpenSlide enables libvips to operate as a bridge between the Aperio SVS image format and Deep Zoom and other International Image Interoperability Framework (IIIF) image formats.

Creating a single Deep Zoom image from an SVS image requires a computer with sufficient software and hardware. When creating multiple Deep Zoom images in parallel on one computer, working memory and local storage can become overloaded, and image conversion will not proceed without intervention. To limit the overloading of the working memory or local storage of the virtual servers for image conversion, the MBRC implemented a distributed computing approach. To create a web gallery of multiple SVS images, each SVS image file is sent to one of several virtual servers in a fleet that allows for server rotation, expansion, and gradual replacement. Three virtual servers operating in parallel can convert 100 SVS image files, distributed as a subset of complete individual SVS images, to Deep Zoom images within a few hours. It is possible to expand this fleet to include many more virtual servers that can accept SVS image files as input, convert the SVS images to Deep Zoom images with libvips, and deliver the Deep Zoom image files to long-term storage. For example, for a gallery with 100 brain sections (i.e., 100 SVS images), 70 images can be converted by the first and third servers (35 each), and 30 by the second server, assuming they are handling sections in rostral-to-caudal order. The reason for the uneven number distribution is that approximately the first 20 rostral images and the last 20 caudal images of a whole brain sliced coronally are generally smaller files based on the physical size of the brain at the rostral and caudal extremes. The intermediate sections are physically larger, represented by logically larger image files, which justifies a variable distribution of work among the servers.

In summary, each SVS image file is located within the eSlideManager storage repository and copied to local storage on the virtual server. SVS image file conversion to Deep Zoom format is performed on local storage. SVS images are distributed in parallel to one of a fleet of virtual servers with libvips and access to the eSlideManager and web gallery server storage repositories. This process produces a Deep Zoom image format file to organize a related set of directories containing pyramid image tiles as JPEG image files.

The dealer protocol, used to distribute batches of SVS image conversion tasks by eSlideManager database ImageID value, is an innovation developed within the MBRC to generate multiple Deep Zoom images in parallel, with the parallelization occurring within and across a fleet of virtual servers. The architecture is scalable by design. Additional storage and working memory capacity would enable more concurrent image conversion tasks per server. Additional servers in the image conversion server fleet would further expand concurrent image conversion capacity. When image conversion is complete, Deep Zoom images are copied from the virtual image conversion

server to the web gallery server storage repository to enable public access. Files related to the image conversion process are removed from the image conversion servers after a successful copy to long-term storage is completed between rounds of image conversion. The system components and relationships of the MBRC image conversion pipeline from Aperio SVS images to Deep Zoom images are shown in **Figure 3**.

2.7.3 Digital output

MBRC web-accessible galleries generally display dozens of Deep Zoom images in a custom-built, responsive, labeled image frame template. The International Image Interoperability Framework (IIIF) is a set of open standards for delivering high-quality, attributed digital objects online at scale [18]. Deep Zoom images can

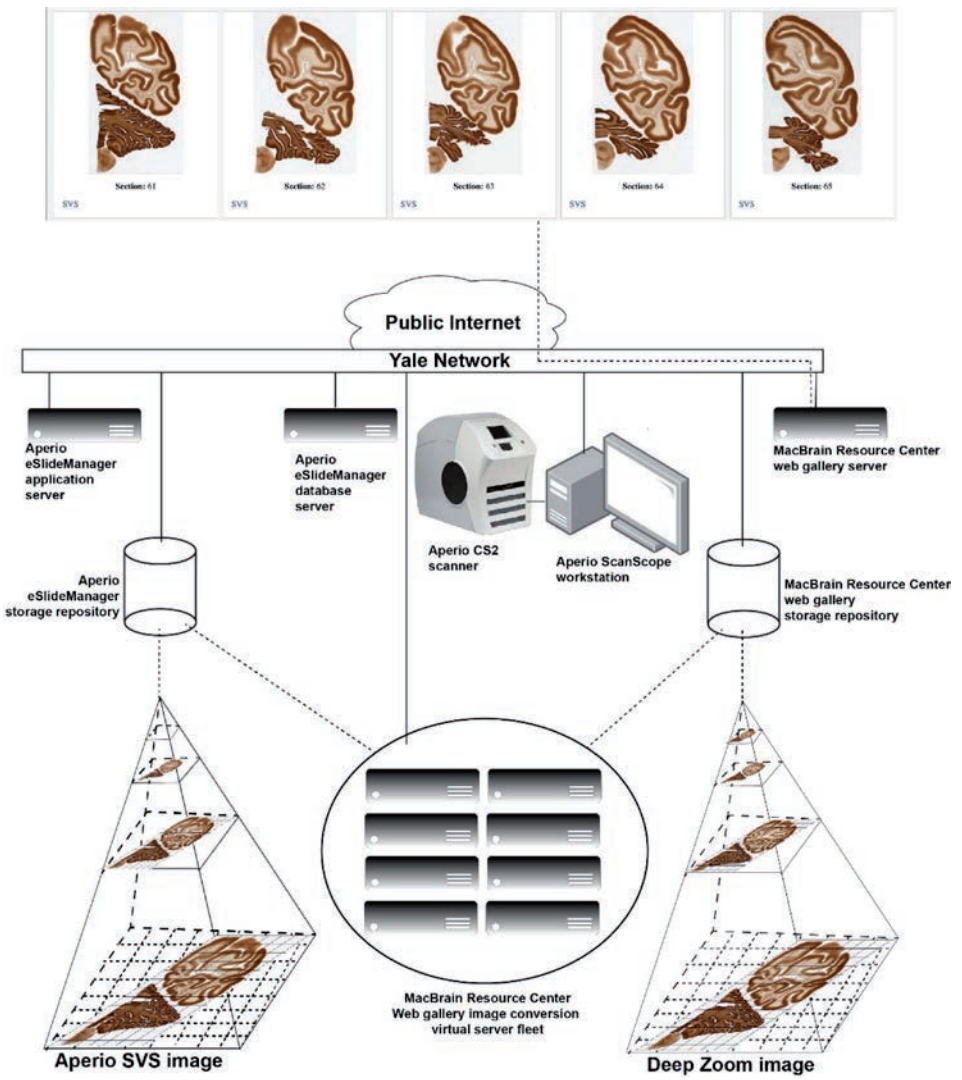

Figure 3.
MacBrain Resource Center (MBRC) image conversion pipeline from Aperio SVS images to Deep Zoom images.

be displayed with IIIF viewers. Seadragon software, the original developer of the Deep Zoom image format, formerly Sand Codex software, was acquired by Microsoft in 2006. OpenSeaDragon is an IIIF image viewer capable of displaying Deep Zoom images [19]. Based on Microsoft's open-source Seadragon Ajax library, MBRC web galleries include JavaScript-based OpenSeaDragon viewer components arranged in a responsive grid to display Deep Zoom, a set of images in each of our web galleries, enabling public access from a range of devices to, at the time of writing, over 30,000 full-resolution, zoomable whole-slide images of over 30 different neuronal and glia markers of 30 NHP brains of diverse age and sex. All images are accessible without plugins or logins, zoomable, and downloadable (**Figure 4**).

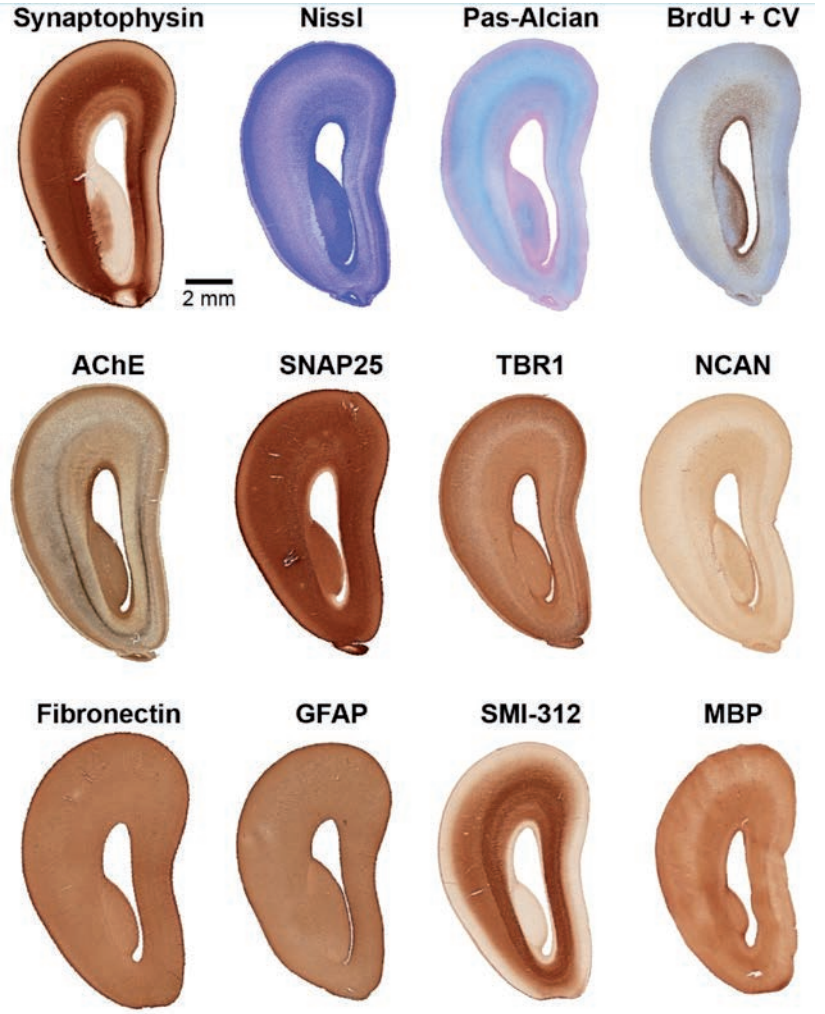

Figure 4.
Histo- and immunohistochemistry of a series of sections stained with different markers. Raw data to study developmental changes. Here, 12 different markers at embryonic age 70 (E70). Gestation in rhesus macaques is on average 165 days. Abbreviations: Periodic Acid-Schiff (PAS)-Alcian blue stain (Pas-Alcian), Bromodeoxyuridine counterstained with cresyl violet (BrdU+CV), acetylcholinesterase (AChE), 25 kDa Synaptosomal-Associated Protein (SNAP25), brain specific T-box transcription factor 1 (TBR1), Neurocan (NCAN), Glial fibrillary acidic protein (GFAP), Pan-axonal Phosphorylated neurofilament protein (SMI-312), and myelin basic protein (MBP). The scale bar applies to all images. Coronal sections of the left hemisphere were cut at 30-micron thickness.

2.8 Leveraging machine learning for NHP brain developmental studies

The rapid expansion of biomedical data, driven by advancements in neuroimaging, electrophysiology, and omics technologies, presents unprecedented opportunities to enhance patient care, deepen disease understanding, and refine treatment strategies [20, 21]. However, conventional analytical approaches, constrained by predefined rules and manual feature selection, frequently struggle to discern the complex patterns embedded within the extensive biomedical datasets [22, 23]. Machine learning (ML) is revolutionizing neuroscience by automating complex data analysis, enabling pattern recognition, and enhancing predictive modeling, thereby increasing research accuracy and efficiency [23–27]. The MBRC, as a primary repository of nonhuman primate (NHP) brain data, presents substantial opportunities for ML investigations into brain development, structural organization, and disease pathology [28]. Integrating ML techniques with MBRC's datasets could yield novel insights into primate brain function and advance neuroscience research.

Despite the wealth of available brain data, conventional statistical methods often fall short of capturing the intricate and nonlinear characteristics inherent in these datasets [29]. Traditional approaches, predicated on predefined assumptions, feature selection, and representative samples, may prove inadequate for extensive datasets. While classical techniques like principal component analysis (PCA) or support vector machines (SVMs) handle specific aspects of high-dimensional data, their scalability for nonlinear interactions remains limited without ML enhancements [30–32]. By contrast, ML uncovers hidden relationships and excels in capturing complex, implicit associations [33]. This capability is particularly evident in deep-learning architectures like convolutional neural networks (CNNs), which have successfully extracted meaningful patterns from immunohistochemistry (IHC) images. For instance, researchers developed LaceDNN, a deep-learning model that improves protein subcellular localization by leveraging label correlations [34]. Similarly, CNNs have shown promise in clinical diagnostics, significantly enhancing the identification of u-serrated patterns in direct immunofluorescence (DIF) microscopy images, a key marker for diagnosing epidermolysis bullosa acquisita [35]. Furthermore, researchers have used CNNs to automatically annotate gene expression patterns in developing mouse brains, outperforming traditional text-derived methods and showcasing the potential of transfer learning in large-scale biological image analysis [36].

The MBRC is pioneering low-cost methods to facilitate brain research by leveraging existing materials. As its collections continue to grow, ML presents a promising opportunity to enhance their utility [37]. ML can streamline key histological analyses within the MBRC, automating tasks such as cell segmentation, protein expression quantification, and tissue classification [38–41]. By reducing human error and significantly accelerating data processing, ML enables more efficient and scalable analysis of vast, high-dimensional datasets [33]. Unlike predefined statistical models, ML discerns complex patterns from data, mapping inputs to outcomes in ways that frequently generalize effectively to unseen cases, contingent on factors such as training data quality, regularization techniques, and model interpretability [29]. This capability renders ML especially valuable for revealing intricate relationships in histopathological data derived from NHP tissue samples. This ability to reveal previously unrecognized cellular and molecular patterns enhances the scientific value of MBRC's broad data archives. By prioritizing prediction and discovery, ML strengthens medical research, amplifying the impact of MBRC's collections in neuroscience.

2.9 Examples of data analysis, combining old and new materials and techniques

Many of the archived materials in the MBRC collections were procured and processed decades ago (e.g., materials in Collections 1 and 5), and a lot of newly processed materials are constantly added (e.g., the expanding Collections 6 and 7). Together, the materials complement each other, therefore avoiding unnecessary repetition of procedures already carried out and unnecessary animal sacrifice. Plenty of *de novo* research has already been published re-mining old materials using newer techniques [6, 7, 9, 42–44]. Below are a few examples of ongoing research taking advantage of these very unique strengths.

2.9.1 Example of de novo research into the genesis of choroid plexus using tritiated thymidine

Despite the importance of the choroid plexus in the production of cerebrospinal fluid (CSF) and the diagnostic importance of CSF itself, information on the kinetics of choroid plexus genesis is extremely scarce. We are currently investigating choroid plexus generation and development by following its evolution in the lateral and 4th ventricles of the developing NHP brain using samples from NHPs exposed to tritiated thymidine (^{3}H-TdR) at different embryonic ages and that were sacrificed at different intervals after injection, including postnatally. Positive cells are being detected by a host of methodologies. The ^{3}H-TdR was developed and visualized by autoradiography in the 1970s. The materials are available in Collection 1 of the MBRC (**Figure 5**).

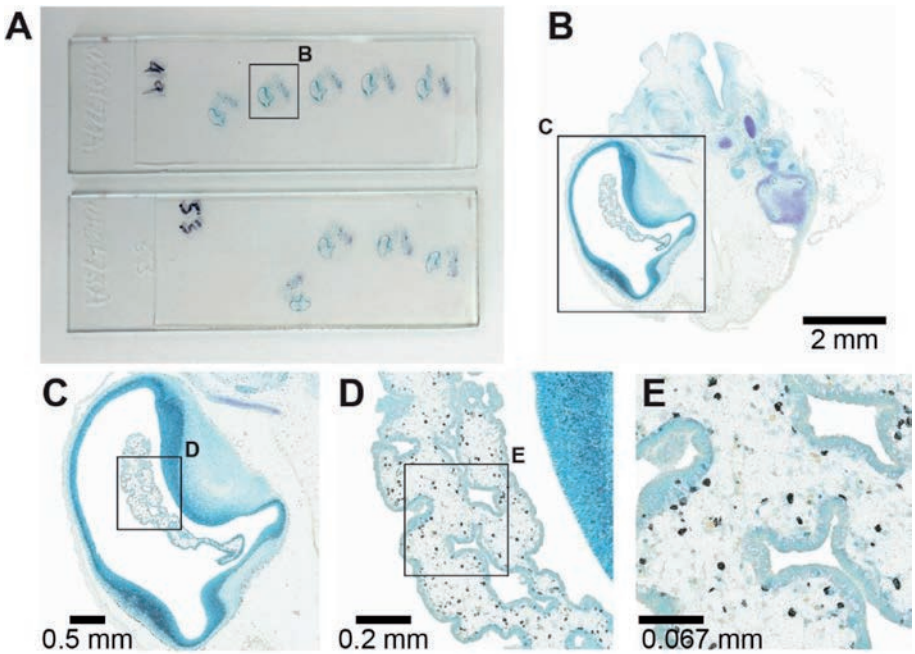

Figure 5.
Two glass slides containing nine coronal sections of embryonic macaque brain tissue (A). The pregnant NHP was injected with ^{3}H-TdR at E41, and the embryo was obtained by C-section and sacrificed an hour later. Sections were processed by autoradiography to reveal labeling of dividing cells at the time of injection and counterstained with Nissl to facilitate observation of landmarks (A–E). This specimen, in MBRC Collection 1, was obtained and processed on June 6, 1973 (hence case number 060673A). Sequentially zoomed-in views of the choroid plexus and surrounding lateral ventricle, highlighting the appearance and strength of silver labeling (B–E).

2.9.2 Example of de novo electron microscopy research of the developing neocortex

Collection 5 of the MBRC consists of thousands of brain tissue segments from NHP embryos of different ages, embedded in epoxy resin and suitable for electron microscopy analysis. **Figure 6** exemplifies materials being analyzed to continue studies of the kinetics of mammalian cerebral cortex expansion. Here, a 55-day-old monkey embryo brain was fixed and embedded in epoxy resin in May 1985. After 40 years of storage, the ultrastructure is perfectly sharp, and the specimen is appropriate for ultrathin sectioning and electron microscopy study, including 3-D reconstruction from serial sections. Some aspects of the long-term research on NHP brain development using Collection 5 of the MBRC have already been published [44–46].

2.9.3 Example of ML at work for the detection of striatal cholinergic interneurons

Figures 7 and **8** illustrate an in-house project that uses histological scans of an E110 female rhesus macaque brain embryo analyzed with an AI-driven tool, a commercially available image analysis software called Aiforia® Create, which enables researchers

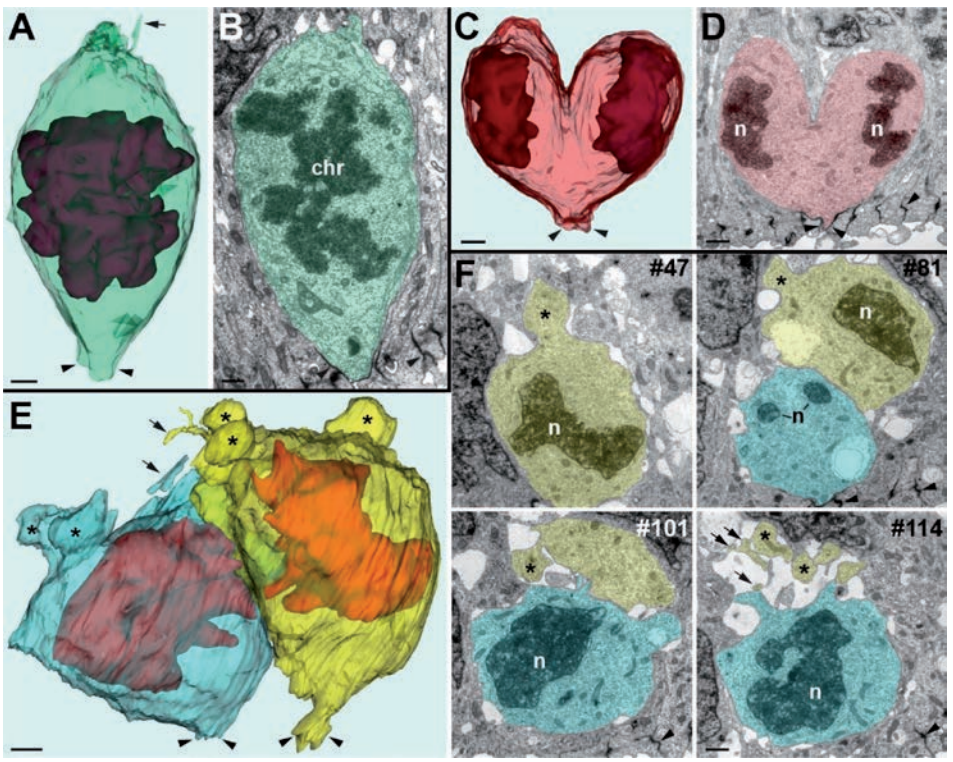

Figure 6.
Electron microscopy 3-D reconstruction from serial ultrathin sections of three mitotic cells in the neocortical ventricular zone of a 55-day-old rhesus macaque embryo. The emergence of condensed chromosomes in the cytoplasm or two opposing daughter nuclei shows that one cell is in metaphase (A, B), the other is in early cytokinesis (C, D), and the third is in late cytokinesis (E, F). Notice that these cells do not have radial processes but thin filopodia (arrows) or short protrusions of the cell membrane (asterisks). The reconstructed cells and their characteristic micrographs are highlighted with corresponding semitransparent colors (green, red, blue, and yellow). Chromosomes (chr) and daughter nuclei (n) are shown inside the cell bodies. The order of the micrographs in the series is indicated by the numbers in (F). The ventricular surface is identifiable with the row of adherens junctions that are indicated with arrowheads. Scale bars are 1 micron.

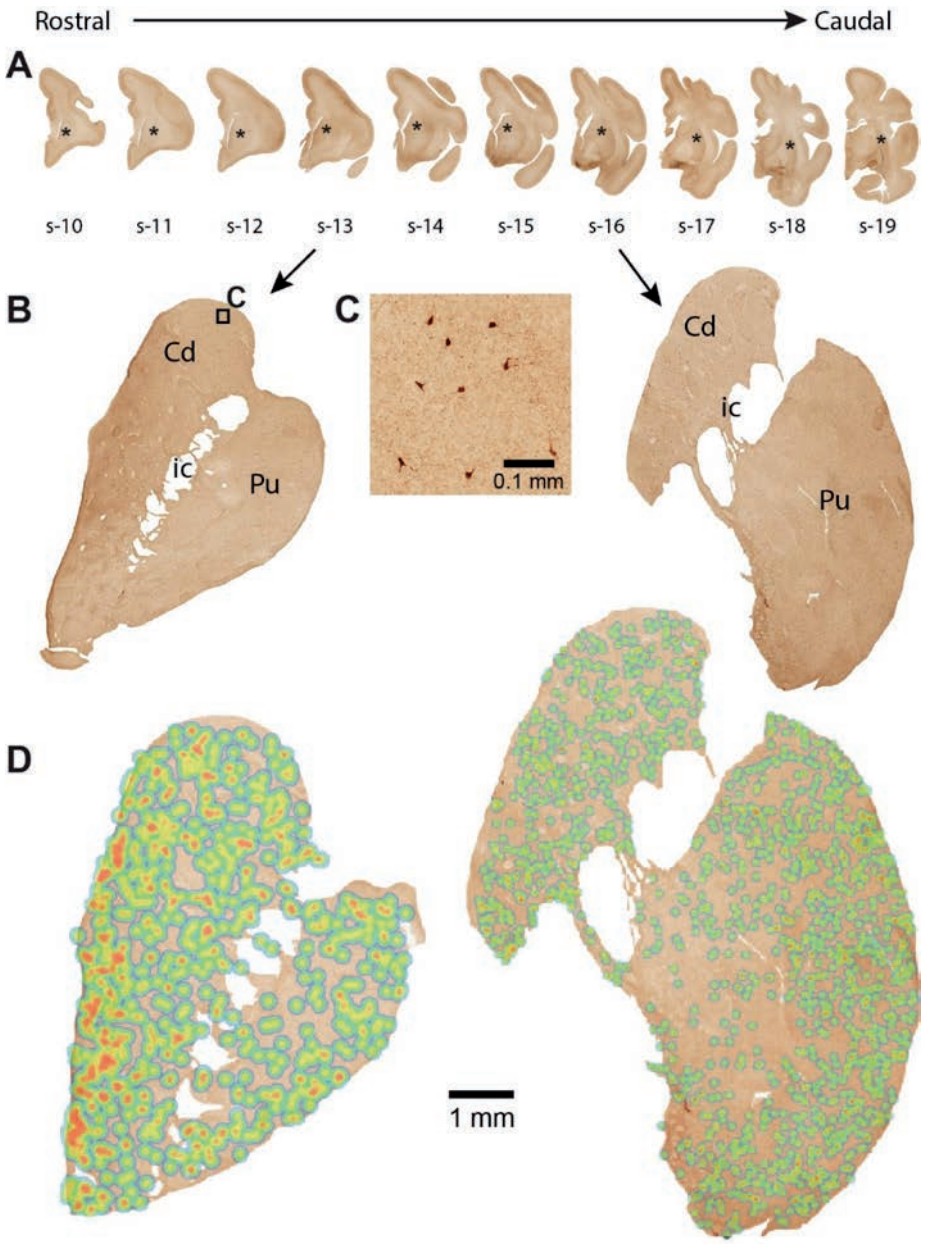

Figure 7.
Process for analyzing striatum in a female rhesus macaque embryo, age: E110. Images of all sections containing striatum are collected for analysis, one hemisphere, rostral-to-caudal (A). The image of each section is edited to contain only the striatum: Caudate and putamen (B). ChAT+ cells within the striatal area are clearly distinguishable at higher magnification (C). After training, Aiforia detects cells, producing a density heatmap (D) and downloadable quantitative data.

to develop ML models tailored to their own research questions. The interface does not require familiarity with written code, only annotations on a set of training images. This technology demonstrates a methodology to mine (and re-mine) existing digital material for new insights into the cytoarchitecture of the NHP brain. Although this example is from a study focused on cholinergic interneurons in the striatum of female

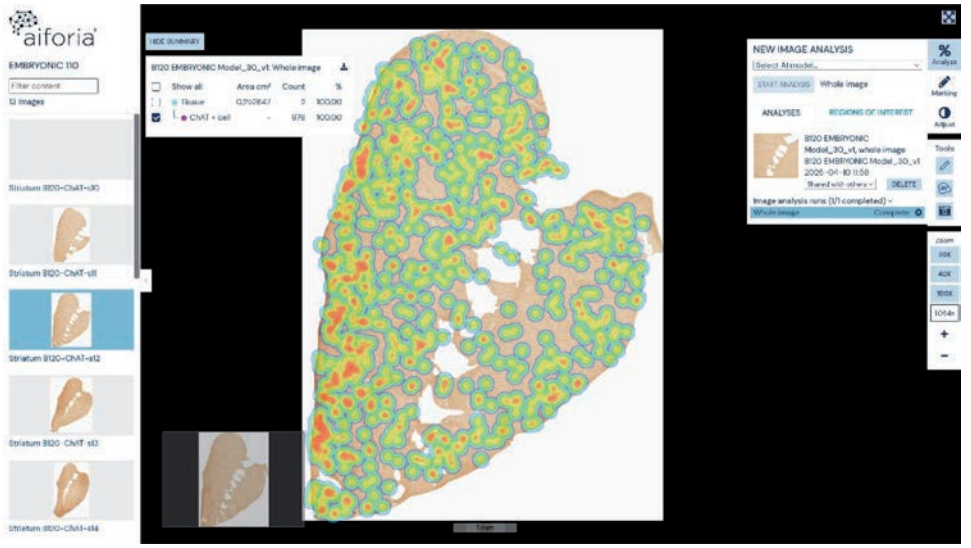

Figure 8.
View of the Aiforia interface after performing image analysis. Results are visualized in a zoomable heat map; quantifications of tissue area and cell counts are summarized on screen, with the option to download a detailed Excel spreadsheet. E110. Serial interactions of ML-based training ensure the algorithm is then able to detect and measure cells very quickly and with a very high level of confidence.

NHPs, the same method can and will be used to detect, count, and measure other features in other regions of the brain and in males. This is an example of tissue from Collection 7 that was specifically processed for ChAT immunohistochemistry, which is then detected and analyzed using ML/AI technology. The materials under study will then be incorporated into Collection 6 and become publicly available.

3. Conclusions

The goal of this chapter has been to raise awareness of the need for and importance of establishing reliable datasets for the study of NHP brain development and to introduce the collections available in the MBRC that can contribute to these efforts. As useful as MRI and other imaging techniques have been, and will continue to be, histology is unsurpassed in studies of cytoarchitecture because of its cellular resolution. We advocate the need to increase the sharing of resources to increase the number of available samples for study. Proper and complete documentation of the samples and of the processing parameters is absolutely necessary to assure usefulness without compromising scientific rigor. We consider the establishment of reliable NHP brain development datasets essential to advance the fields of neurodevelopment and neuroscience, and we have started to process tissue to populate the already established Collection 6 of the MBRC with such materials.

Acknowledgements

This work is supported by NIH grant MH113257 (AD). Studies that contributed embryonic brain hemispheres and blocks to the MBRC are supported largely by DA023999 (PR).

Conflict of interest

The authors declare no conflict of interest.

Notes/thanks/other declarations

We thank Ashley Ake, Esmeralda Alejo, Mari Elliott, and Mikhail Rodov for scanning materials for analysis.

Author details

Alvaro Duque*, Phil Barello, Lucy Greene, Emma Burke, Valeria Mendoza-Silva, Yury M. Morozov and Pasko Rakic
Department of Neuroscience, MacBrain Resource Center, Yale University School of Medicine, New Haven, CT, USA

*Address all correspondence to: alvaro.duque@yale.edu

References

[1] Mitsunaga F, Ueiwa M, Kamanaka Y, Morimoto M, Nakamura S. Fetal sex determination of macaque monkeys by a nested PCR using maternal plasma. Experimental Animals. 2010;**59**(2):255-260

[2] Junaković A, Kopić J, Duque A, Rakic P, Krsnik Ž, Kostović I. Laminar dynamics of deep projection neurons and mode of subplate formation are hallmarks of histogenetic subdivisions of the human cingulate cortex before onset of arealization. Brain Structure & Function. 20230102nd ed. 2023;**228**(2):613-633

[3] Li H, Duque A, Rakic P. Origin and development of the claustrum in rhesus macaque. Proceedings of the National Academy of Sciences of the United States of America. 20230705th ed. 2023;**120**(28):e2220918120

[4] Micali N, Ma S, Li M, Kim SK, Mato-Blanco X, Sindhu SK, et al. Molecular programs of regional specification and neural stem cell fate progression in macaque telencephalon. Science. 20231013th ed. 2023;**382**(6667):eadf3786

[5] Rash BG, Arellano JI, Duque A, Rakic P. Role of intracortical neuropil growth in the gyrification of the primate cerebral cortex. Proceedings of the National Academy of Sciences of the United States of America. 20221227th ed. 2023;**120**(1):e2210967120

[6] Sakharkar M, Rockland KS, Duque A. Complex neurochemical microstructure of the stria terminalis in infant and adult macaque monkey. Frontiers in Neuroanatomy. 20220525th ed. 2022;**16**:891608

[7] McDonald AJ, Duque A. Specific neuronal subpopulations in the amygdala of macaque monkeys express high levels of nonphosphorylated neurofilaments. Brain Research. 20211224th ed. 2022;**1777**:147767

[8] Rash BG, Duque A, Morozov YM, Arellano JI, Micali N, Rakic P. Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum. Proceedings of the National Academy of Sciences of the United States of America. 20190320th ed. 2019;**116**(14):7089-7094

[9] Duque A, Krsnik Z, Kostović I, Rakic P. Secondary expansion of the transient subplate zone in the developing cerebrum of human and nonhuman primates. Proceedings of the National Academy of Sciences of the United States of America. 20160808th ed. 2016;**113**(35):9892-9897

[10] Duque A, Gazula VR, Kaczmarek LK. Expression of Kv1.3 potassium channels regulates density of cortical interneurons. Developmental Neurobiology. 20130911th ed. 2013;**73**(11):841-855

[11] Duque A, Tepper JM, Detari L, Ascoli GA, Zaborszky L. Morphological characterization of electrophysiologically and immunohistochemically identified basal forebrain cholinergic and neuropeptide Y-containing neurons. Brain Structure & Function. 20070602nd ed. 2007;**212**(1):55-73

[12] Duque A, Balatoni B, Detari L, Zaborszky L. EEG correlation of the discharge properties of identified neurons in the basal forebrain. Journal of Neurophysiology. 2000;**84**(3):1627-1635

[13] Hsu SM, Raine L, Fanger H. Use of avidin-biotin-peroxidase complex (ABC)

in immunoperoxidase techniques: A comparison between ABC and unlabeled antibody (PAP) procedures. The Journal of Histochemistry and Cytochemistry. 1981;**29**(4):577-580

[14] Hsu SM, Raine L. Protein A, avidin, and biotin in immunohistochemistry. The Journal of Histochemistry and Cytochemistry. 1981;**29**(11):1349-1353

[15] Aperio Technologies Inc. Digital Slides and Third-Party Data Interchange [Internet]. Vista, CA, United States; 2008. Available from: http://www.aperio.com/documents/api/Aperio_Digital_Slides_and_Third-party_data_interchange.pdf [Accessed: March 6, 2025]

[16] Balkan A. OpenSlide [Internet]. 2013. Available from: https://openslide.org/ [Accessed: March 6, 2025]

[17] Microsoft. Deep Zoom File Format Overview [Internet]. 2011. Available from: https://learn.microsoft.com/en-us/previous-versions/windows/silverlight/dotnet-windows-silverlight/cc645077(v=vs.95)?redirectedfrom=MSDN [Accessed: March 6, 2025]

[18] Framework III. How it Works [Internet]. Available from: https://iiif.io/get-started/how-iiif-works/ [Accessed: March 6, 2025]

[19] Project O. DZI Tile Source [Internet]. Available from: https://openseadragon.github.io/examples/tilesource-dzi/ [Accessed: March 6, 2025]

[20] Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. The Lancet Oncology. 2019;**20**(5):e262-e273

[21] Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. BMC Medical Informatics and Decision Making. 2018;**18**(1):139

[22] Huang X, Li Z, Zhang M, Gao S. Fusing hand-crafted and deep-learning features in a convolutional neural network model to identify prostate cancer in pathology images. Frontiers in Oncology. 20220927th ed. 2022;**12**:994950

[23] Zhang J, Xia Y, Xie Y, Fulham MJ, Feng DD. Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. IEEE Journal of Biomedical and Health Informatics. 2018;**22**:1521-1530

[24] Cascarano A, Mur-Petit J, Hernández-González J, Camacho M, Eadie N, Gkontra P, et al. Machine and deep learning for longitudinal biomedical data: A review of methods and applications. Artificial Intelligence Review. 2023;**56**:1711-1771

[25] Seo H, Badiei Khuzani M, Vasudevan V, Huang C, Ren H, Xiao R, et al. Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. Medical Physics. 2020;**47**(5):e148-e167

[26] Delli Priscoli M, Memmolo P, Ciaparrone G, Bianco V, Merola F, Miccio L, et al. Neuroblastoma cells classification through learning approaches by direct analysis of digital holograms. IEEE Journal of Selected Topics in Quantum Electronics. 2021;**27**:1-9

[27] Buchlak QD, Esmaili N, Leveque JCA, Farrokhi FR, Bennett CC, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: An artificial intelligence augmented systematic review. Neurosurgical Review. 2019;**43**:1235-1253

[28] Duque A. Reducing Non-Human Primate Sacrifice for Brain Research in the Virtual Era [Internet]. Open Access Government; 2023. Available from: https://www.openaccessgovernment.org/reducing-non-human-primate-sacrifice-for-brain-research-in-the-virtual-era/170646/

[29] Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment. Medicina (Kaunas). 20200908th ed. 8 Sep 2020;**56**(9):1-10

[30] Morishita T. Time-dependent principal component analysis: A unified approach to high-dimensional data reduction using adiabatic dynamics. The Journal of Chemical Physics. 2021;**155**(13):134114

[31] Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. AICHE Journal. 1991;**37**:233-243

[32] Li B, Artemiou A, Li L. Principal support vector machines for linear and nonlinear sufficient dimension reduction. Annals of Statistics. 2011;**39**:3182-3210

[33] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: A practical introduction. BMC Medical Research Methodology. 20190319th ed. 2019;**19**(1):64

[34] Hu JX, Yang Y, Xu YY, Shen H. Incorporating label correlations into deep neural networks to classify protein subcellular location patterns in immunohistochemistry images. Proteins: Structure. 2021;**90**:493-503

[35] Shi C, Meijer JM, Azzopardi G, Diercks GFH, Guo J, Petkov N. Use of convolutional neural networks for the detection of u-serrated patterns in direct immunofluorescence images to facilitate the diagnosis of epidermolysis bullosa acquisita. The American Journal of Pathology. 20210628th ed. 2021;**191**(9):1520-1525

[36] Zeng T, Li R, Mukkamala R, Ye J, Ji S. Deep convolutional neural networks for annotating gene expression patterns in the mouse brain. BMC Bioinformatics. 2015;**16**(1):147

[37] Duque A. Fostering nonhuman primate (NHP) brain research without animal sacrifice. Open Access Government. 2023;**38**:186-186. DOI: 10.56367/OAG-038-10792

[38] Kan A. Machine learning applications in cell image analysis. Immunology and Cell Biology. 20170315th ed. 2017;**95**(6):525-530

[39] Durkee MS, Abraham R, Clark MR, Giger ML. Artificial intelligence and cellular segmentation in tissue microscopy images. The American Journal of Pathology. 2021;**191**(10):1693-1701

[40] Claeys T, Menu M, Bouwmeester R, Gevaert K, Martens L. Machine learning on large-scale proteomics data identifies tissue and cell-type specific proteins. Journal of Proteome Research. 2022;**22**:1181-1192

[41] Rączkowski Ł, Możejko M, Zambonelli J, Szczurek E. ARA: Accurate, reliable and active histopathological image classification framework with Bayesian deep learning. Scientific Reports. 2019;**9**(1):1-12

[42] Spadory T, Duque A, Selemon L. Spatial-temporal topography in neurogenesis of the macaque thalamus.

Brain Structure and Function. 2022;**227**(5):1673-1682

[43] Ahmed B, Duque A, Rakic P, Molnár Z. Correlation between the number of interstitial neurons of the white matter and number of neurons within cortical layers: Histological analyses in postnatal macaque. The Journal of Comparative Neurology. 2024;**532**(6):e25626

[44] Morozov YM, Rakic P. Lateral expansion of the mammalian cerebral cortex is related to anchorage of centrosomes in apical neural progenitors. Cerebral Cortex. 3 Jul 2024;**34**(7):1-9

[45] Morozov YM, Mackie K, Rakic P. Cannabinoid type 1 receptor is undetectable in rodent and primate cerebral neural stem cells but participates in radial neuronal migration. International Journal of Molecular Sciences. 20201117th ed. 17 Nov 2020;**21**(22):1-19

[46] Arellano JI, Morozov YM, Micali N, Rakic P. Radial glial cells: New views on old questions. Neurochemical Research. 20210316th ed. 2021;**46**(10):2512-2524