Statistical signatures of learning in the olfactory bulb and a new approach to image stack segmentation

Corentin Guérinot

SORBONNE UNIVERSITÉ

École doctorale Cerveau, Cognition, Comportement ED 158

Unit Genes, Synapses and Cognition, CNRS UMR-3571

Laboratory Perception & Memory

Laboratory of Decision and Bayesian Computation

Neuroscience Department, Computational Biology Department, Institut Pasteur

Statistical signatures of learning in the olfactory bulb and a new approach to image stack segmentation

By **Corentin GUÉRINOT**

Under the supervision of

Pierre-Marie LLEDO and **Jean-Baptiste MASSON**

Abstract

With the development of genetically encoded fluorescent calcium indicators, the recording of numerous neurons with 2-photon microscopy in an awake, behaving animal has become a well-established neuroscience technique. This tool has several advantages for making behavioral correlates with neuronal activity to provide further understanding of the circuits required for learning and memory. The vast amount and the heterogeneity of the data produced entail difficulties in identifying true correlations between neural activity and behavior. It is pivotal in particular to make sure that identified effects are not spurious correlations occurring due to the high-dimensional nature of the inference. These challenges are compounded when comparing between animals.

The main focus and driving objective of this thesis has been to identify statistical signatures of learning in the neural activity of olfactory bulb granule neuron dendrites recorded using 2-photon calcium imaging as the animal learned an operant conditioning task. This is a non-trivial inference task since the granule cell population is sparsely labeled and the imaged population differs completely between different animals. To tackle this problem, we employed self-supervised learning to project dendrite activity time-series to a common low-dimensional latent space across animals. Performing supervised classification using these highly compressed latent representations of neural activity as input allowed us to identify that the neural activity of olfactory bulb granule cells encode highly predictive features for the odor presented to the animal and for whether it makes the correct behavioral choice (*i.e.* gets the reward). We furthermore identified a robust signature of the animal's learning encoded in the neural signal, though with a lower predictive accuracy than for odor and reward.

Identifying dendrite structures in microscopy images is a hard image segmentation problem. Addressing this problem naturally led to the developments presented in the second part of this thesis which combines human-in-the-loop analysis with one-shot learning in a virtual reality environment. This project became another main axis of the thesis beyond our initial neuronal microscopy analysis. Our focus was to address data analysis when classic machine learning is not possible due either to the lack of massive data or to the complexity in interacting with the data. We combined virtual reality, user annotation, and cloud based one-shot learning to exemplify we could accelerate data treatment. We demonstrated the efficiency of the approach for both microscopy and medical applications.

Preface

Table of Contents

Abbreviations

AAV	Adeno-associated virus
AE	Autoencoder
ANN	Artificial neural network
ANOVA	Analysis of variance
AR	Augmented reality
CART	Classification and regression trees (algorithm)
CNS	Central Nervous System
CR	Correct rejection
CT	Computed tomography (scan)
DCX	Doublecortin
EGFP	Enhanced green fluorescent protein
EGTA	Egtazic acid
ELBO	Evidence lower bound
EM	Electron microscopy
EPL	External plexiform layer
FA	False alarm (incorrect detection)
FoV	Field of view
GABA	Gamma-aminobutyric acid
GCaMP	Genetically encoded calcium indicator
GC	Granule cell
GCL	Granule cell layer
GL	Glomerular layer
GLRT	Generalized likelihood ratio test
HELA	Immortal cell line from Herietta Lacks

Hi-C	High-resolution chromosome conformation capture
HITL	Human-in-the-loop
HSCT	Hematopoietic stem cell transplantation
M/T	Mitral and tufted (cell)
MCL	Mitral cell layer
MD	Molecular dynamics
MMD	Maximum mean discrepancy
MRI	Magnetic resonance imaging
MSE	Mean squared error
NaN	Not a number
OB	Olfactory bulb
ONL	Outer nerve layer
OOB	Out-of-bag (samples / error)
OSN	Olfactory sensory neuron
PCA	Principal component analysis
PFA	Paraformaldehyde
ReLU	Rectified linear unit
RF	Random forest
RMS	Rostral migratory stream
RMSE	Root mean squared error
ROI	Region of interest
SEM	Standard error on the mean
SSL	Self-supervised learning
VAE	Variational autoencoder
VR	Virtual reality
xAI	Explainable artificial intelligence

Part 1: Statistical signatures of learning

1 Introduction

1.1 Calcium indicators for measuring neuronal activity

Neurons are the primary cells of the central nervous system responsible for acquiring, integrating, and relaying signals with the neuronal network functioning to interpret inputs for processing further output to drive numerous tasks, including sensory processing, memory, learning, cognition, and many physiological roles. Early studies demonstrated how neurons utilize neurotransmitter receptors to receive input, which causes changes in membrane voltage that is then either shunted or relayed to other neurons within the network[1]. Traditional methods directly measure this voltage change, utilizing various forms of patch-clamp recordings of individual neurons with a glass electrode[2], field recordings of isolated units in a network, electroencephalography of broad network changes, or recently, multi-electrode arrays for more precise broad network detection[3].

Calcium (Ca^{2+}) is a key ion used by neurons as a depolarizing signal driven by Ca^{2+} influx via calcium channels on the plasma membrane, along with Ca^{2+} release from internal stores[4]. Based on this essential role of Ca^{2+}, organic fluorescent calcium indicators were developed in the early 1980's by synthesizing calcium chelators (e.g., EGTA) with fluorescent dyes made of the binding of Ca^{2+} enhancing their fluorescent efficiency and thereby allowing detection of neuronal activity with fluorescence microscopy methods[5]. These dyes were used extensively in cultured neurons and brain slice preparations which allowed expanding single-cell recording into imaging broad network activity measurements. With the advent of 2-photon *in vivo* imaging, bulk loading of calcium indicator underneath a coverglass cranial window of the brain in an intact mouse was

performed, which provided a natural, complete network without cell culture or brain slicing artifacts[6].

In vivo imaging with organic calcium, indicators had some distinct disadvantages including non-specific labeling of various cells within the infusion zone, not restricted to neurons and including astrocytes, microglia, oligo-precursor cells etc., concentration gradients causing irregular dye concentration in cells, loss of dye loading across days limiting time-course imaging to short durations (hours), and the acute dye injection with window surgery causing artifacts due to the inflammatory response[6]. With the cloning and enhancement of a green fluorescent protein (EGFP) derived from jellyfish[7] it was possible to use genetic methods to label cells for tracking and visualizing their complex structure, which started a quantitative revolution in the biological sciences. A genetically encoded green fluorescent Ca^{2+} indicator (GCaMP) was cloned by combining calmodulin with EGFP where the calmodulin Ca^{2+} binding domain undergoes a conformational change that increases EGFP fluorescence, thus allowing the detection of fluorescent changes that correlate with neuronal activity[8].

Further refinement in protein engineering to induce multiple single-point mutations for changing the GCaMP structure with high throughput screening led to the development of numerous brighter and more robust forms of GCaMP, providing better signal-to-noise and faster transition kinetics between Ca^{2+} bound/unbound states[9]. Knock-in mice for GCaMP were also developed where expression is induced by crossing with cre-recombinase neuron (and subtype) specific lines or GCaMP driven by various Thy-1 promoters produced for specific labeling of various neuronal populations[10]. Additionally, numerous viral expression vectors were developed under neuron-specific promoter control for more convenient viral labeling of neuronal populations in wild type and transgenic mice[9].

1.2 In vivo 2-photon imaging

The advent of the 2-photon microscope in 1990 was one of the greatest biological tool discoveries in the last 30 years since it allows exploring organisms in their intact and natural state[11]. Utilizing a pulsed infrared laser attached to a microscope with a highly

sensitive detector, one can image large brain volumes at high resolution with relatively little photobleaching of the fluorophore, great depth penetration and with minimal disturbance of cells[11]. For brain imaging, a craniotomy with a cover glass window or thinned skull preparation is performed in transgenic mice expressing a fluorescent protein or labeled with viral techniques. Early imaging studies focused on structural changes in the brain, uncovering microglia dynamics[12], and neuron dendritic[13,14] and spine[15–17] structural plasticity. Coupling 2-photon imaging with labeling of neurons with GCaMP has allowed awake imaging of networks of neurons to understand correlations between neuronal activity and animal behavior[18–20].

1.3 Olfactory bulb and granule cell function

The olfactory bulb (OB) is a brain structure dedicated to the initial processing of olfactory sensory information that is directly received from the nasal olfactory epithelium[21]. Olfactory sensory neurons line the epithelium where they extend cilia that contain olfactory sensory receptors, with the receptor type defining the chemical detection specificity of the neurons while also being the cue for proper targeting to the connectivity of the OB[22]. The axons of olfactory sensory neurons project through the cribriform plate of the skull and target specific glomeruli within the glomerular layer of the OB (Fig. 1). The glomeruli are a dense spherical neuropil structure that contains the synapses of the olfactory sensory neuron axons with the dendritic tufts of the principal output neurons of the OB, mitral and tufted cells (M/T cells). Each glomerulus receives a singular input from only olfactory sensory neurons that express the same olfactory receptor type, thus providing specificity of input signal and convergence for greater detection thresholds.

Granule cells (GCs) are the most abundant neuron population in the OB, with these GABAergic interneurons providing the primary inhibitory drive to M/T cells through their unique dendro-dendritic synapses[21]. GCs do not have an axon, instead they form reciprocal synapses on the smooth dendritic shafts of M/T cells with their large dendritic spines. GCs receive glutamatergic input from M/T cells while providing reciprocal GABAergic release onto M/T cells for local and lateral inhibition[21]. This unique synapse

has been shown to be essential for enhancing odor discrimination[23] while being associated with odor memory[15] and olfactory learning[24,25].

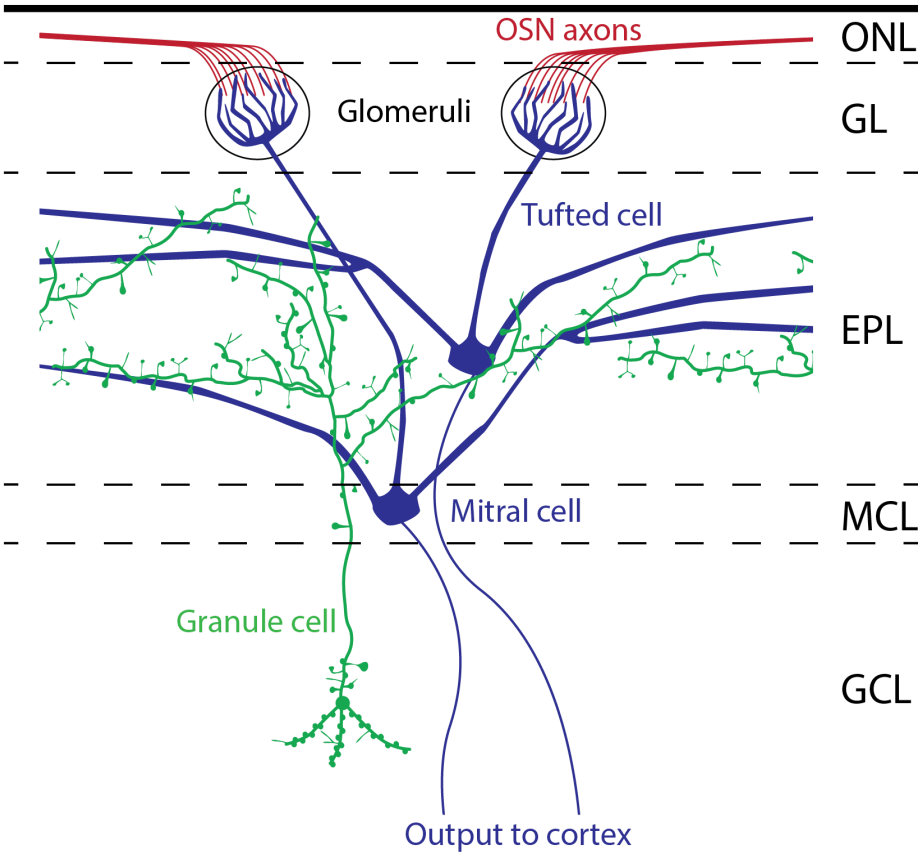


Figure 1: Organization of olfactory bulb (OB) circuit. Diagram showing the OB anatomy with its major segregated layers: outer nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL) and granule cell layer (GCL). Input from olfactory sensory neuron (OSN) axons in the ONL converge to glomeruli of the same olfactory receptor type where they make synapses with M/T cells in their dendritic tufts confined to the glomerulus. Granule cells in the GCL project their apical dendrites in the EPL where they make dendro-dendritic synapses with M/T cells, receiving glutamatergic input while reciprocally releasing GABA locally and laterally. M/T cells integrate the inputs and fire action potentials in their axons which synapse with neurons in the piriform cortex.

Learning causes changes in brain circuits to adapt and be able to predict a response when presented with the same or similar stimulus[28]. This ability to generalize provides an advantage to an organism for having rapid, correct responses when placed in a natural context such as in food scavenging and threat detection, with the OB being the

first point of olfactory learning. Due to the high dimensionality of odors and with ~1,000 odor receptors in mice[29], a high degree of processing needs to occur to provide generalized output to the cortex for rapid behavioral responses[30]. GCs are highly plastic neurons that mediate strong control over M/T cells to greatly shape the direct sensory input, along with neurogenesis which has been shown to be key in olfactory perceptual learning[31]. For operant conditioning, where an animal is presented with two separate odors and learns to associate one odor with a reward, stimulating the adult-born GC population accelerated learning[26], and interestingly, just activating GCs, without an odorant, was sufficient to cause reward-oriented behavior[32]. Taken together, the involvement of GCs in learning is clear, yet the causal, or even correlative relationship between GC activity and performance in learning is still undetermined. To find a signature of learning that can predict learning across animals is the main focus of this part of the thesis.

1.4 Experimental limitations to detecting learning signatures in olfactory bulb granule cells

The mechanisms underlying how GC activity changes for learning an olfactory task are still unclear, largely due to experimental limitations. Traditional slice physiology, either electrophysiology or spinning disk Ca^{2+} imaging, is unsuitable because numerous connections are lost in the preparation causing an incomplete network, a natural odor stimulus cannot be applied, and importantly, behavior effects on the network can only be compared across animals of learned *versus* control groups after sacrificing. Modern *in vivo* patching techniques are currently unsuitable because GCs are some of the smallest cells in the CNS. Previous work has used immediate early genes to determine changes in activity with learning[32], however, it provided a representation of GCs that were active during the task, without any temporal indication of activity and could not observe the evolution of activity with learning. We chose 2-photon *in vivo* Ca^{2+} imaging of an operant conditioning task since it allows measuring the activity of hundreds of GCs during the pre-odor, odor presentation, reward presentation, and the post-odor period for finding correlates with specific GC activities and odor/reward predictions.

Previous studies exploring the role of GCs in the OB also utilized Ca^{2+} imaging. In an early study, the authors used 2-photon imaging to explore the effect of anesthesia on OB circuits[33]. They imaged M/T cells and GC somatic activity and demonstrated GCs to be highly silenced by anesthesia. They also showed a gradual decrease in M/T cell activity with daily exposure to an odorant, implying GC inhibition as a perceptual learning signal. In another study using optical fiber photometry, where a fiber optic probe is implanted in the OB of bulk loaded GCs, enhanced separation of odor responses in GCs was correlated with an operant conditioning task[34]. Both these studies showed GCs having a strong role in the OB, but also introduced further questions. The generalized change across the population of GCs was measured in these studies, since the GC somas were imaged in the first study between days without tracking the same individual GC and the photometry of the second study measured the average fluorescent change of thousands of GCs beneath the optical fiber. It was our goal to provide sparse enough labeling of GCs to have individual regions of interest that could be tracked across days and weeks, which introduced numerous new technical hurdles in the following statistical analysis.

1.5 Predicting learning from neural activity in the olfactory bulb

The problem we are trying to address is a non-trivial statistical learning problem. The dendrites we image are only a small subset of the dendrites present in the OB, and from one mouse to another we have no fine control over the number of marked dendrites nor the local density of dendrites. Thus from one mouse to another, we will image different subsets and, in the same mouse, we do not have access to the complete dendrite signals. In addition, it is possible to have several signals belonging to the same dendrites in different parts of the image without having the means to test it directly. Furthermore, some dendrites are only active for certain odors, and therefore the signal may be absent for several weeks and only appear during certain days. Finally, optical imperfections and minor registration variations result in slightly out-of-focus signals whose selection as an area of interest is sometimes uncertain.

These different constraints had a strong influence on the nature of the procedure we developed for odor identification and learning. In order to test for a statistical signature of learning and to make the procedure transferable to all mice, it was necessary

to construct a cumulative representation of the recorded neural signal. The variability in the number of neurons imaged, in their temporal activities and in the coding of the different sensory signals makes the construction of this common representation at the scale of the neuron population and even at the scale of individual neurons impossible. However, a common point of all the experiments is the local activities of the dendrites. Here, local simply means the activity of a small portion of neurons whose activity is non-zero for the duration of the experiment, in other words a region of interest (ROI). Thus, we decided to reduce the total neuronal activity to the activity of ROIs, defined as spatial regions wherein the signal varied at least once during the total duration of the experiments.

The reduction of a time-series of images to a set of time-series of ROI activities allows for a common representation to be built over time and for all animals. Under this configuration, detecting statistical traces of information will be done by predicting the odors and their proportion by the activity of the ROIs as well as predicting the learning scores from these same activities. Thus, our approach is placed in the framework of statistical learning. Specifically, we learn a low-dimensional latent representation of the neural activity using an autoencoder architecture, where the latent space is the layer of the autoencoder of smallest size and constitutes the bottleneck of the autoencoder network which forces a compressed representation of the data retaining only the most important features for reconstructing the input data. We have chosen an over-constrained approach, where the latent space we are going to use is of a very small size, for two reasons. The small size of the latent space allows us to explore this space, to characterize its properties and to relate them to biological activity. The second is to protect against overfitting. A major risk when using a powerful machine learning approach is that it will find spurious results that only emerge due to the randomness and the finite size of the samples, *i.e.* it may *overfit*. By highly compressing the data, *i.e.* massively reducing the dimension of its representation in the latent space, the model cannot perfectly regenerate the experimental data but we reduce its ability to overfit.

All subsequent analysis was based on these vectors summarizing the activity of the ROIs. However, even if the dimensional reduction of the ROIs is substantial, the possibility of overfitting is still present as each mouse was characterized by several hundred ROIs.

We are clearly far from the orders of magnitude belonging to big data, but there are enough dimensions that overfitting might be an issue. In order to ensure that the elements learned are linked to reliable information, we trained the same neural networks on data where we permuted the labels for the associated odors and actions. These randomized data keep the same structure of the neural activity time-series but without any correlation with the odor presented and the action of the mouse. If the performance of the network was identical for experimental and randomized data, this would mean that it is the network that is able to associate any input with an output and not a proof of learning. Conversely, if the performance is significantly higher on the experimental data, there is indeed information that has been derived from the neural activity.

2 Methods

2.1 Calcium imaging data acquisition and preprocessing

2.1.1 Neuron labeling and cranial window procedure

All experimental procedures were performed following approved guidelines as regulated by the local ethics committee (Comité d'éthique en experimentation animale no. 89). For all surgical procedures, male mice (12 weeks, n = 16) were anesthetized with ketamine/xylazine and body temperature was maintained with a rectal thermometer feedback heating pad. Adult-born cells were labeled with a lentiviral tdTomato (tdTom) cre-recombinase construct by intracerebral injecting into the rostral migratory stream at the following coordinates: anterior = +3.3 mm, lateral = ±0.8 mm, ventral = +2.9 mm (a). After 3 weeks, to allow the labeled neural progenitors to migrate to the OB, a craniotomy over the OB was made and Adeno-associated virus (AAV) (2/1) encoding a floxed genetically encoded calcium indicator (GCaMP6f, University of Pennsylvania Viral Vector Core) under the synapsin promoter was injected 5 times, every 100 mm from 500 to 900 mm. Thus tdTom-Cre^{+} cells specifically expressed GCaMP6f. A 1.4 x 3.0 mm custom cut cover glass (University of Wisconsin-Madison Center for Applied Microelectronics) was inserted into the craniotomy, a stainless steel head bar was glued to the skull (cyanoacrylate) and the window edge and skull were covered with dental cement (Superbond) as previously described[19]. After the surgical procedure, the animals were returned to their cage with free access to food and water in a room at ambient temperature (21C, 40-60% humidity) with 14-hour light and 10-hour dark cycle.

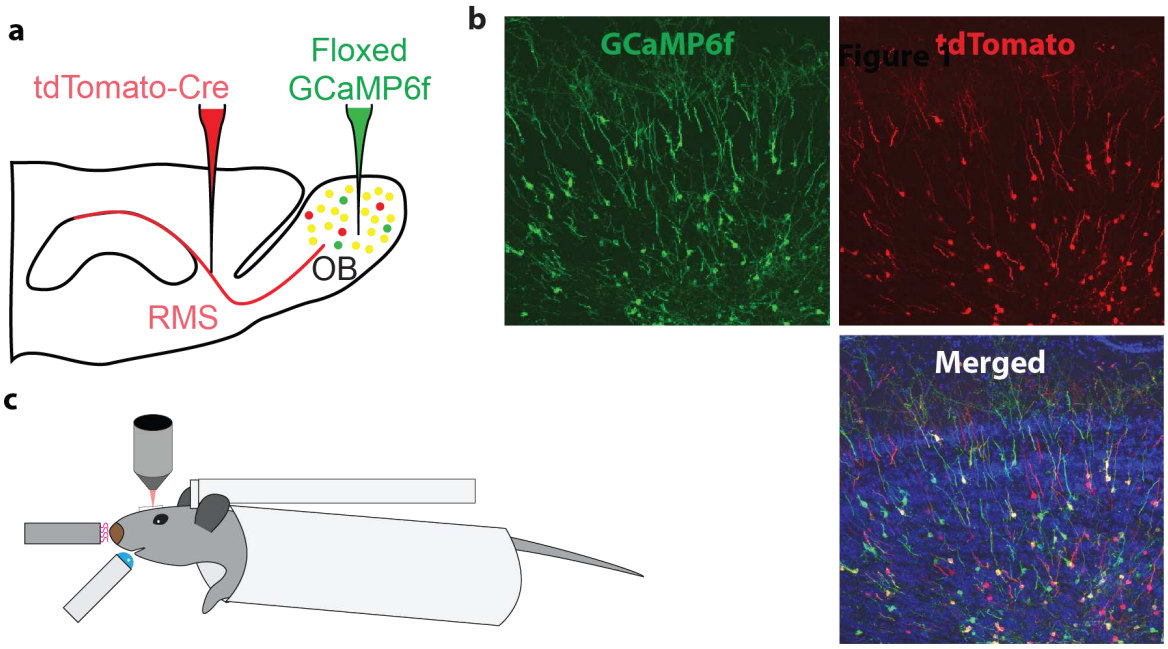


Figure 2: Olfactory learning experimental setup. (a) Labeling strategy where lentivirus expressing tdTomato-Cre is injected into the rostral migratory stream (RMS) 3 weeks prior to injection of AAV expressing floxed-GCaMP6f into the OB. (b) Confocal images of OB 12 weeks after labeling showing GCaMP6f (green), tdTomato (red) cells labeled with numerous cells co-expressing (yellow, merged). (c) Imaging setup with mouse on head-fixed stage in a Falcon tube, odor delivery to the nose, water tube for lick detection and water delivery beneath the mouth and cranial window with objective for imaging.

2.1.2 In vivo imaging and behavioral task

Mice were trained for awake head-mounted imaging for 1 week prior to imaging with handling and familiarization with the mounting stage and an open tube for positioning the mouse (50 ml Falcon tube). For imaging, the mouse was mounted by securing the head bar to the mounting stage with the mouse in a tube for support and comfort. This stage was secured to the microscope stage and a water tube and odor delivery tube were positioned to the mouth and nose, respectively (c). Ultrasound gel was used for immersion and a 16x Nikon 0.8 NA objective on an Ultima Investigator microscope (Bruker) was used for imaging with a 2-photon laser (Mai Tai, Spectra-Physics) tuned to 950 nm. A 2-channel olfactometer was used to randomly deliver odors, detect licking, deliver rewards and record the mouse performance, as previously described[9]. After finding the field of view from the previous imaging session, a

z-stack was acquired 20 mm above and below the visually identified imaging plane. Using custom software (Matlab), a 2-D cross-correlation was performed with each image *versus* the previous session 2-D reference image, the stack was registered to the reference image and the cross-correlation was performed again with the z-stack image of the highest correlation z-distance used to set the proper focus of the microscope. This technique allowed imaging of the same imaging plane over multiple weeks. Images of both the GCaMP6f and tdTomato signal were acquired at an X-Y resolution of 1.66 mm/pixel using a resonant galvanometer imaging at 15 Hz with 2-frame averaging (b).

For the behavioral go-nogo task odorants (Sigma-Aldrich) were diluted 10% in mineral oil with valeraldehyde *versus* cineole for the easy task and 60% ethyl-butyrate + 40% amyl-acetate versus 40% ethyl-butyrate + 60% amyl-acetate the mice were used for the difficult task. The final odor concentration was 1% with a 1:10 dilution with delivery air. The mice were water-deprived to 80-85% of their body weight to be receptive to a water reward. For a behavioral trial, image acquisition started 8 seconds prior to odor delivery, which lasted 2 s, and imaging continued for 10s for a total of 20 seconds per behavioral trial. Twenty trials consisted of a block and the behavioral results were recorded as Hit: S^+ odor + lick, Miss: S^+ odor + no lick, correct rejection (CR): S^- odor + no lick, false alarm (FA): S^- odor + lick and a score was calculated ((hits + CRs)/20 X 100). Airflow (humidified air) was constantly delivered, 2 seconds prior to odor delivery the odorant was mixed with delivery air in a circuit diverted to vacuum. At odor delivery, a final valve switched to the odor delivery circuit. If the mouse performed 3 licks of the water tube during the 2 seconds S^+ odor presentation, the mouse received a 3-ml water reward immediately after odor delivery. The same reward delivery occurred when the mouse did not lick for the S^- odor.

2.1.3 Ca^{2+} neuronal activity preprocessing

Images were converted to 3D TIFFs for each block containing 20 trials, a reference image was made taking the median of all images in the tdTomato channel and 2-D cross-correlation values were recorded for each tdTomato image versus the reference image. A manual threshold was assigned to each block based on the histogram of

cross-correlation values and imaging frames below this threshold were removed in the GCaMP6f channel for excluding frames that had movement artifacts.

For registering between blocks, each block TdTomado signal was average projected, affine registration was performed using FIJI ("Multistackreg", available at http://bradbusse.net/downloads.html) while recording the registration values and these values were then assigned to all GCaMP6f images within the respective block. Manual ROIs were traced on a z-stack excluding movement frames in the GCaMP6f signal for any dendritic segment that showed activity for all trials across all days.

There are limitations in the efficiency of automated signal extraction for dendrites signals. Most of them are designed for cell body signal extraction[15]. In the present study, relevant signals can stem from a small portion of dendrites with most of the corresponding cell structure being out of the imaging plane. For this reason, regions of interest (ROIs) were manually selected for each field of view (FoV). The ROIs were refined with PCA analysis[26], keeping the first principle component to more accurately detect activity-associated pixels and adjust for small shifts between sessions. On average, 303 portions of dendrites were extracted and their normalized fluorescence was computed. Baseline fluorescence was calculated using an autoregressive process on the 2 seconds preceding odor presentation. Neuronal activity was indicated as the normalized change in GCaMP6f fluorescence: (F – F0)/F0.

Missing frames due to movement in the Z-axis were replaced with NaN in order to ensure the temporal integrity of each trial. Trials were excluded i) if missing values represent more than 25% of the total trial duration, or ii) if there is a minimum of 30 consecutive missing values in the trial. Missing values are often located during odor and reward presentation, both provoking animal movement and thus Z-plane shift. The above procedure leads to the exclusion of 4.5% of all trials, representing on average 71 trials per animal. In order to pursue further analysis, the remaining missing values are replaced through linear interpolation.

After this first quality control, we filtered out some ROIs exhibiting abnormal normalized fluorescence signals. These ROIs showed peaks of abnormally high amplitude unrelated to calcium transients given their fast rise and decay in around 0.15 second, far

faster than GCaMP6f typical kinetics[27]. Such ROIs are usually located on the edges of the FoV and present extremely low baseline fluorescence. Empirically, maximum normalized fluorescence ranges from 10 to 16, hence we decided to exclude ROIs exceeding a threshold of 16, resulting in an average loss of 6.6% of all ROIs. At the end of the full process, 11.2% of total data was removed due to one of the anomalies mentioned before.

2.2 Machine learning techniques

The recent progress in computation combined with the advances in statistical theories have enabled the emergence of machine learning in all quantitative fields[26,28]. Machine learning refers to the process of conception, analysis, and implementation of methods enabling the machine – *i.e.* the computer – to learn and build predictions based on data.

In this work, we employ random forests (RFs), which are an efficient and robust ensemble learning technique, for supervised classification[29]. For learning a common low-dimensional representation of neural activity across trials and animals, we employ autoencoders (AEs) and variational autoencoders (VAEs), which are a generative self-supervised learning approach.

2.2.1 Random Forests

Random forests (RFs) are an efficient supervised learning approach. They were developed by Breiman et al.[29,30] and employ an ensemble of decision trees to make predictions on unseen using rules learned from a labeled training dataset. RFs can be used for two different purposes:

- *classification*, *i.e.* assigning a class label to each sample;
- *regression*, *i.e.* predicting a continuous output variable for each sample.

In this work, RFs have been used as a binary classifier to predict behavioral outcomes during olfactory learning experiments, but also in 3D semantic segmentation in Part II of

the thesis. The RF algorithm for classification consists in the construction of an ensemble of decision trees. The final prediction is made by the majority vote in the ensemble of trees, and the estimation of the performance on unseen data is made through out-of-bag error assessment or cross-validation.

a. Decision Trees

Decision trees are fundamental tools in data mining, enabling the quick sorting of large datasets through a succession of binary decisions. As a predictive model, it has become a widely used supervised learning technique, taking several predictor values as an input and outputting a class label (or continuous value in the case of regression trees).

In terms of architecture, a decision tree is made of nodes and splits to the data, initialized with the entire training set in the first node. According to the CART (Classification and Regression Trees) algorithm[30], the dataset is segmented into two child nodes using a predictor variable following a greedy procedure. The split is selected to maximize a measure of the *purity* of the child nodes, which measures how homogeneous the resulting classes are in terms of the data labels. The construction is performed so forth until reaching terminal nodes. From the latter, no further split can be made insofar as it would not increase the purity. Predictions are then made on the content of these terminal nodes, following an adapted decision path.

The choice in the splits is paramount when growing a decision tree, it is made using specific impurity metrics. The two most popular metrics are the entropy and the Gini index G; the CART algorithm is based on the latter. In a node m, representing N_m observations, let $p_{m,k}$ denote the proportion of observations in class k,

$$p_{m,k} = \frac{1}{N} \sum_{x_i \in R_m} I(y_i = k), \tag{1}$$

then the Gini index is defined as

$$G = \sum_{k=1}^{K} p_{m,k} \left(1 - p_{m,k}\right). \tag{2}$$

The Gini index measures in each node the probability to misclassify a randomly labeled element from the set[31]. Splits are then made according to their decrease in Gini impurity at each node.

Decision trees exhibit several advantages: they are fast to grow and do not require lots of computation power in predictions. However, decision trees have high variance. Indeed, given the hierarchical process of splitting, an early error can easily propagate and lead to an important instability. As a consequence, the risk is that the tree detects noise inherent to the training set rather than significant patterns paramount for generalization: this phenomenon is called overfitting. A pruning step is usually performed after growing the tree in order to minimize this effect.

To overcome these challenges, several statisticians successively proposed and optimized the concept of RFs, obtained by bootstrap aggregation[26] (bagging) of decision trees, and which show higher generalization accuracy[26].

The concept of bagging lies in the averaging of many noisy yet approximately unbiased models[26] – here decision trees – thus reducing the associated variance and leading to a more stable final predictor. For prediction of the class of a new data point, the majority vote of the ensemble of trees is used, *i.e.* the most probable class obtained in the RF.

b. Parameter tuning and characteristics

Breiman demonstrated that RF accuracy depends on two characteristics[29]:

- the correlation between trees in the RF – reduced by the bootstrapping procedures and the random variable selection at each split – the lower the better;
- the strength of each tree in the RF – accuracy of individual trees – the higher the better.

The former decreases monotonically with the number of features selected for each split, m, while the latter increases monotonically with m. One can thus adjust m to find the optimal trade-off between them. Another hyperparameter to tune is B, the number of

trees in the RF. Generally, the higher B is the better as it decreases the variance of the RF without increasing its bias; it is usually picked so that the accuracy stabilizes. Finally, the minimal number of datapoints in a node, n_{min}, directly controls the depth of each individual tree in the RF. It thus impacts their bias and variance – smaller n_{min} leads to lower bias and higher variance – and that of the RF.

RFs exhibit excellent accuracy among competitive algorithms, capable of effectively handling missing data and outliers using proximities. The multiplication of decorrelated trees implies an important decrease in variance while keeping the bias low, hence the drastic limitation of overfitting paramount when generalization is sought[30]. The bagging process makes it possible to estimate prediction accuracy without cross-validation through the use of out-of-bag samples.

c. Out-of-bag samples

Theoretically, one can demonstrate that a bootstrap sample -- *i.e.* N samples randomly drawn with replacement from an N-sized dataset -- overlaps about two-third of the original dataset, for N big enough. Each tree in the RF is grown using a different bootstrap sample from the original dataset, letting thus about one-third of the samples left out and not used to build each individual tree. Those are called out-of-bag (OOB) samples, and by definition have not been seen yet by the tree, therefore they can be used as a classification test set yielding an unbiased estimate for the performance.

Let $(x, y) = (\{x_1, x_2, ..., x_N\}, \{y_1, y_2, ..., y_N\})$ denote our data set of input variables and labels, respectively. One looks at the majority vote of the trees for which x is OOB and compares the so-obtained class to the true one. Repeating this process over the whole dataset, the probability that the predicted class is not equal to the true class averaged over all classes is called the OOB error estimate. The main advantage of this method is that no cross-validation or separated test set is needed to assess the predictive accuracy unbiasedly, allowing RF to perform well on relatively small datasets.

Furthermore, OOB samples can be used to assess each variable's importance, *i.e.* their prediction strength. To evaluate the importance of variable p, one should for each

tree pass down the OOB samples and record the prediction accuracy. Then the values for the variable p are randomly shuffled among the OOB samples and the corresponding accuracy is again computed; the difference in accuracy obtained for each tree is finally averaged over the RF. Variable importance is usually expressed in percentage, by construction they add up to 100%, and they stand as an unbiased estimate of prediction driving strength of each variable, allowing comparison amongst them.

2.2.2 Autoencoders

Experiments are characterized by a wide variety of visible dendrites in the FoV between mice. Hence, the number of dendrites, the number of ROIs and the global activity in the FoV are not constant between experiments. We devised our analysis scheme to address ROI activity. Hence, while the number of ROIs vary between experiments, the encoding strategy of ROIs is unique to all the mice. We trained an autoencoder (AE) to encode ROI activity while enforcing a strong dimensionality reduction to compress a given trial activity into a limited number of hidden variables to ensure the robustness of the analysis.

a. Principles

AEs are artificial neural networks (ANNs) for unsupervised learning, trained to reconstruct as precisely as possible their input. They were first introduced in 1986 by Hinton *et al*[32], in part of their pioneering work on gradient backpropagation for training ANNs. With the advent of powerful computation capacities in the 2000's, deep architectures for ANNs have emerged[33], making AEs a very popular and accessible tool. They can be used for various purposes:

- ➢ data compression or dimensionality reduction;
- ➢ signal processing, for denoising or anomaly detection;
- ➢ feature extractor for transfer learning.

In this work, AEs have virtually served all the above goals: denoising neuronal activity traces, and exploring latent space representation in relation to animal behavior, also as a feature extractor for semantic segmentation.

AEs are trained to compress and reconstruct as precisely as possible their input, basically mimicking the identity function. Their architecture includes two distinct and complementary parts: an encoder and a decoder. They are usually symmetrical with respect to the middle layer, constituting a bottleneck where the information is compressed. The idea is that the bottleneck forces the AE to learn a space-efficient latent representation of the data that retains the important features of the data but filters out the noise. The resulting latent space (the middle layer) provides a new representation of the data in lower dimension, usually more accessible for further analysis or graphical illustration.

Formally, the encoder e encodes a low-dimensional latent representation of each (high-dimensional) data point, $h = e(x)$, while the decoder d outputs a reconstruction of the input from its latent position, $\hat{x} = d(e(x))$. The encoder and decoder are each parameterized by an ANN which are trained together to minimize a loss between $\hat{x}$ and x (see below).

b. Training autoencoders

As an example, fully connected AEs are built similarly as a multilayer perceptron, consisting of successive layers of neurons densely connected. The architecture may vary according to the dataset of interest, for instance through the addition of convolution layers for image processing. Autoencoders offer the advantage of being trained in an unsupervised manner, implying they do not require labeled data and can be used directly.

Fitting an AE consists in finding an optimal set of weights leading to the most accurate reconstruction of the input. It is done through the minimization of an objective function, called the loss function. It is determined by the user according to the task to perform. The loss function depends on the type of data: often mean squared error (MSE) are chosen for continuous values and cross-entropy for categorical ones. MSE, here in this

study, is calculated as the average of the squared differences between the predicted outputs and the input values,

$$L\left(\hat{x}, x\right) = \frac{1}{N}\sum_{i=1}^{N} ||\hat{x}_i - x_i||_2^2 \,. \tag{3}$$

The minimization of the objective function is done through an optimization algorithm called gradient descent. Several versions for this algorithm have been developed in the last decade, making training step fast and computationally effective: stochastic gradient descent, RMSprop, Adam[34] to name but a few. At each optimization step, the parameters of the model are adjusted using backpropagation.

2.2.3 Variational autoencoders

Because learning is a temporal process, its neural signature is harder to detect than neural encodings of the odors detected and/or actions triggered. To further regularize the latent space representation of neural activity time-series we employ a Bayesian variational extension of AEs, the variational autoencoder (VAE)[35,36]. In VAEs, the statistical inference problem of variables (or model inference) is replaced by an optimization problem on an objective function. Unlike basic AEs, inputs are here encoded as probability distributions over latent space. This makes VAEs generative models and helps prevent overfitting.

a. Principles

While the initial goal is to offer a new representation in a low-dimensional subspace, there is no evidence that the structure of this subspace will conserve initial information. The lack of regularity in the latent space makes AE an imperfect tool for new content generation[26]. Two properties are here pivotal:

- ➢ Continuity: close points in the latent space should give similar decoded outputs
- ➢ Completeness: any sampled point from the latent space should produce meaningful output

In order to satisfy at best these properties and limit overfitting, VAEs have been introduced[37]. While they share common features with AE such lossy reconstruction in output and similar architectures, some differences are introduced:

- ➢ Encoding: instead of outputting a discrete representation h, a distribution is encoded in the latent space. The encoder then has two outputs: the expectation of the latent variable, E(z), and its associated variance V(z). Several options can be used here, however a normal distribution is usually preferred.
- ➢ Sampling: before decoding, a sampler is required to generate the latent variable z in the distribution parameterized by the encoder.
- ➢ Decoding: finally, the sampled variable z can be decoded into $\hat{x}$ the same way as in an AE.

VAE training is done through loss function minimization, over x. The loss function can be written as follows:

$$L(\theta, \phi) = -\mathbb{E}\left[\log\left(p_\phi(x|z)\right)\right] + D_{\mathrm{KL}}\left(q_\theta(z|x)||p_\phi(z)\right). \quad (4)$$

This loss is composed here of a reconstruction term, $E[\log(p_\phi(x|z)]$, as found in AEs, and a regularization term, $D_{KL}(q_\theta(z|x)\,||\,p_\phi(z|x))$, which ensures appropriate properties in the latent space. This additional term is called the Kullback-Leibler (KL) divergence, it measures the difference between the latent distributions of the data points $x_i \in x$ and the prior distribution $p_\phi(z)$, which is chosen to be a standard normal distribution for convenience. Minimizing both terms together enables accurate reconstruction and encoded latent distributions that are close to Gaussian distributions.

b. Variational interpretation

Let $q_\theta(z|x)$ denote the inference model for stochastic encoding and $p_\phi(x|z)$ denote the generative model for probabilistic decoding. As both input x and encoding z are random variables, we can write the joint probability defined by the VAE and express it as likelihood and prior:

$$p_\phi(x, z) = p_\phi(x|z)p_\phi(z). \quad (5)$$

Using Bayes theorem, we have:

$$p_\phi(x|z) = \frac{p_\phi(z|x)p_\phi(x)}{p_\phi(z)}. \quad (6)$$

The true posterior $p_\phi(z|x)$ being intractable, we use variational inference to approximate it, and instead compute the variational approximation $q_\theta(z|x)$ that minimizes the Kullback-Liebler distance $KL(q_\theta(z|x)||p_\phi(z|x))$. We can rewrite this divergence as:

$$D_{\mathrm{KL}}\left(q_\theta(z|x)||p_\phi(z|x)\right) = \log\left(p_\phi(x)\right) + D_{\mathrm{KL}}\left(q_\theta(z|x)||p_\phi(z)\right) - \mathbb{E}\left[\log\left(p_\phi(x|z)\right)\right]. \quad (7)$$

Since the KL divergence is non-negative, we have that:

$$\log\left(p_\phi(x)\right) \geq \mathbb{E}\left[\log\left(p_\phi(x|z)\right)\right] - D_{\mathrm{KL}}\left(q_\theta(z|x)||p_\phi(z)\right). \quad (8)$$

The right-hand side in Eq. (8) is called the Evidence Lower Bound (ELBO[26] in short) since it provides a lower bound on the marginal likelihood $p_\phi(x)$, termed the *evidence* in Bayesian statistics. Equation (8) implies that minimization of the initial KL divergence between the true posterior $p_\phi(z|x)$ and its approximation $q_\theta(z|x)$ is equivalent to the maximization of the ELBO.

c. "Reparameterization trick"

Minimization of the VAE loss function is performed by maximization of the ELBO, allowing us to perform approximate inference. It avoids the intractable computation of the KL divergence between the approximate and exact posteriors. Assuming a normal distribution for $q_\theta(z|x)$ with unknown expectation μ and variance σ, the objective function can be optimized using gradient descent.

The sampler step between the encoder and the decoder does not allow however to perform backpropagation. In order to allow synaptic weight update throughout the VAE, it is possible to reparametrize the sampling for z by introducing $\epsilon \sim N(0, 1)$ and let

$$z = \mathbb{E}[z] + \epsilon \, \mathbb{V}[z]^{1/2}. \quad (9)$$

This function depends deterministically on the parameters, making training using gradient descent possible. Reparametrizing z using eq. (9) is known as the *reparametrization trick*[37].

2.2.4 Model training

a. Training dataset

For each calcium trace, we created a vector corresponding to the normalized fluorescence of a ROI during a whole trial. In order to obtain comparable data across ROIs from different animals, we divide each animal's activity by its global maximum. It results in vectors with values ranging from -0.25 to 1 with a baseline preserved at 0.

A vast majority of individual traces represent only rapid and low amplitude fluctuations around the baseline. In order to build a statistically relevant dataset, we oversampled ROIs where significant activity could be detected. We defined an empirical threshold for relevant activity where at least 7 timepoints above 3 standard deviations should be detected to be considered to be an active ROIs. The obtained subset of informative traces represents around 5 % of the original data.

b. Autoencoder training and hyperparameter tuning

Model hyperparameters have been tuned using a grid search cross-validation process[38], namely batch size and number of epochs, training algorithm, learning rate, weight initialization method, input and output activation function, dropout regularization, architecture and dimension of the latent space. For the autoencoder, the mean squared error (MSE) was used as a loss function, which can be interpreted as a reconstruction

error between the input trace and its reconstructed output. For the variational autoencoder, we used the regularized loss function defined earlier.

For both the AE and VAE models, we opted for a symmetrical architecture with 4 densely connected layers of 300, 100, 25 and 4 neurons. Input and output layers have a linear activation function, whereas the remaining have a 'ReLU' activation function; all weights were initialized using 'Glorot normal' initializer[39]. We use extreme compression leaving only a vector of size 4 to capture all possible dynamics of the ROIs.

Training was performed on 225 000 samples and tested on a large portion of the dataset containing also 225 000 samples. Both the training and the test samples had an equal proportion of randomly selected informative and non-informative traces across all animals. The models were fitted on 128-sample batches for 100 epochs, using the 'Adam' optimizer[34] with a learning rate of 0.001.

c. Random forest training and hyperparameter tuning

We trained random forest classifiers (RFCs) taking as input either the latent vectors encoded by the AE or the VAE to predict different behavioral outcomes: i) the odor presented during a trial, ii) whether it was rewarded or not, and iii) the training level of the animal during the trial. Letting N be the number of ROIs in a given FoV, each trial corresponding to N calcium signal time-series, these are then encoded in a $4N$ vector in the latent space. Unlike the AE and VAE, which were fitted on the whole dataset, the predictions of the RFC are animal specific. However, only the latent representation is used to build the prediction, which ensures that different animals can be compared. We used a 5-fold cross-validation procedure to tune the hyperparameters of the RFCs.

In order to ensure that a higher-than-random prediction accuracy was not obtained solely by chance or due to overfitting, a RFC model was trained for each animal with shuffled labels. We then used a generalized likelihood ratio test (GLRT) to assess the statistical significance of the prediction accuracy for learning based on the VAE encoding of the neural activity traces by comparing to the prediction accuracy of the RFCs trained on shuffled data.

3 Results

3.1 *In vivo* imaging of operant conditioning task revealed activity changes after learning but difficult to ascertain between animals

To achieve sparse labeling of GCs we injected lentivirus expressing tdTomato and recombinase (tdTomato-Cre) into the rostral migratory stream (RMS) where adult-born progenitor cells reside. The labeled cells migrate to the OB via the RMS where they differentiate into GCs, among other cell types. Three weeks after lentivirus injection a craniotomy was performed over the OB, AAV expressing a floxed version of the genetically encoded calcium indicator GCaMP6f under the synapsin promoter was injected into the OB.

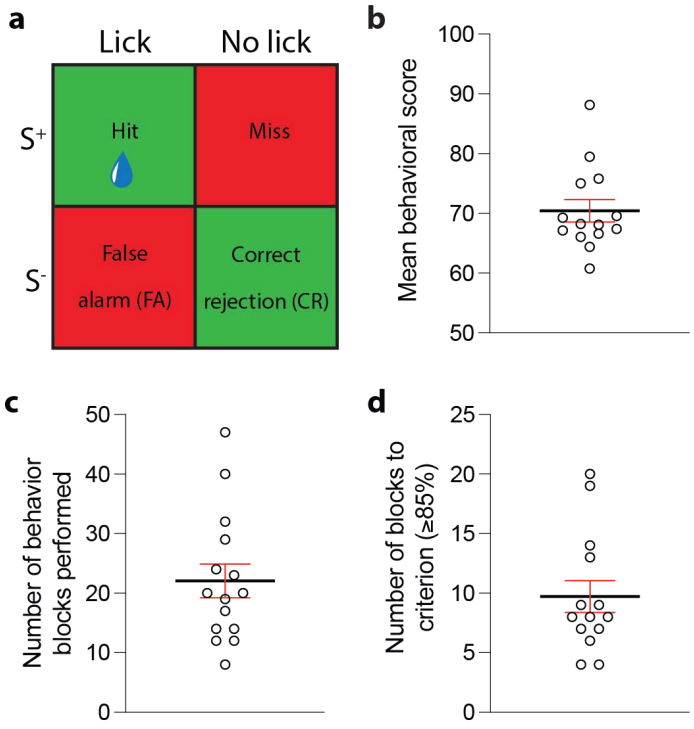


Figure 3: Operant conditioning task and behavior. (a) Operant conditioning task showing correct (green) and incorrect (red) behavior. For all "hits", the animal received a water reward. (b) Mean behavioral score of all animals in the easy task for all blocks (n = 15 animals, mean ± SEM). (c) Total number of blocks each animal performed in the easy behavioral task (n = 15 animals, mean ± SEM). (d) Number of blocks until animal reached behavioral criterion (≥85%) for each animal (n = 15 animals, mean ± SEM).

Thus, expression of GCaMP6f was selectively driven in GCs labeled with tdTomato-cre. A glass cranial window was utilized for long-term imaging over multiple time points. A mixture of expressions was observed in GCs based on the differential expression of each fluorescent protein.

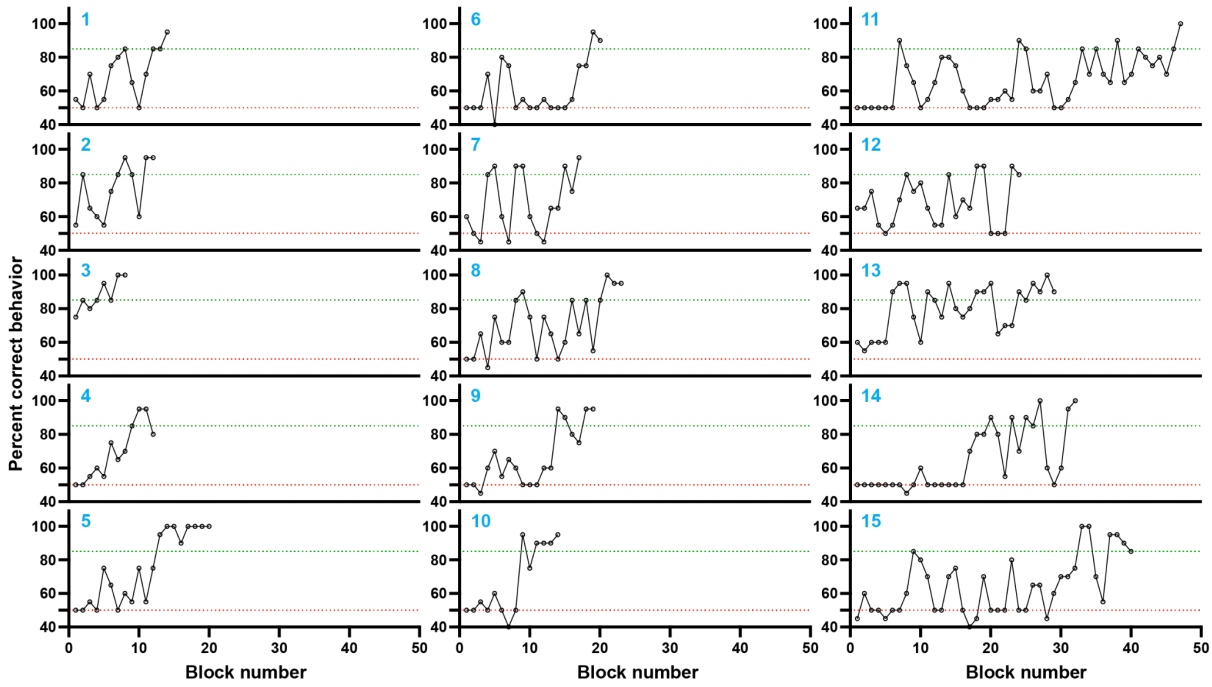


Figure 4: High variability in learning across animals. Learning traces of each animal showing percent correct behavior across blocks with each block consisting of 20 trials. Random, no learning (50%, red dotted line) and criterion for learning (≥85%, green dotted line) levels are shown.

The mice were water-deprived, and after training the animal for head-fixed position it was placed on a stage with a computer-controlled olfactometer providing odor delivery to the nose and a water tube with lick detection and for delivering water reward. The animals were first exposed to the S^+ and S^- odorants without reward provided (passive odor). For the operant conditioning task , the correct behavior was for the mouse to lick during the S^+ (Hit) presentation and to not lick during the S^- (Correct rejection, 'CR') odor presentation. The mouse was delivered a water reward only for "Hits". Incorrect behaviors were when not licking for the S^+ (Miss) and licking for the S^- (False alarm, 'FA'). The mice gradually learned the task, yet there was high variability in the learning rate, the persistence of learning, and final learning achievement.

The same single imaging plane in the external plexiform layer was imaged (at 15 Hz) for all experiments to capture GC dendritic fluorescent activity changes during the task. After X-Y registration and filtering for out-of-Z-plane movement, ROIs were selected on all the GC dendritic structures that showed fluorescent activity (see Fig. 5a. Each trial lasted 20 seconds with a 2 seconds odor exposure, and numerous ROIs showed odor-evoked responses as shown in Fig. 5b. Individual analysis of mice showed a small population of GC ROIs activated by odor exposure, which was also highly variable between animals, probably due to the sparse labeling in a sparsely connected network.

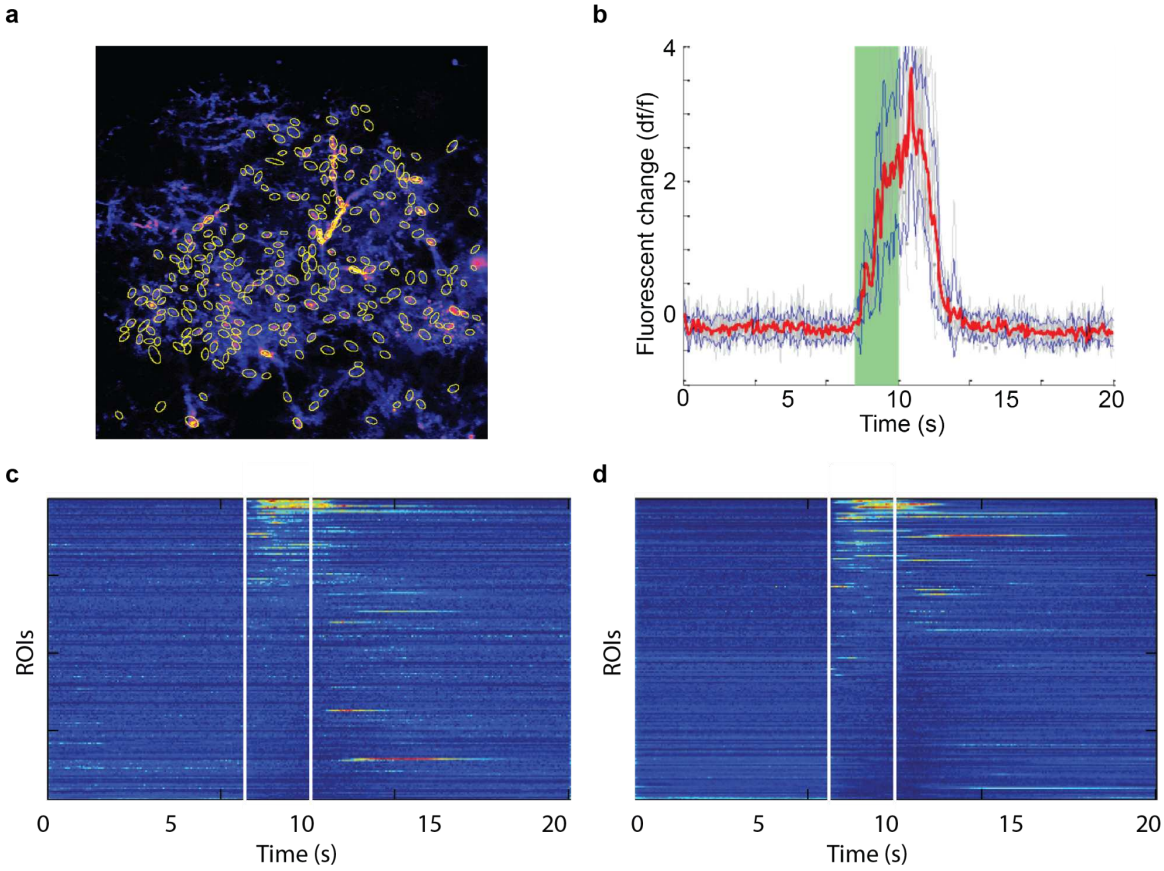


Figure 5: Ca2+ imaging ROI selection and GC activity. (a) Example 2-photon in vivo image of GCaMP6f time projected of a field of view with multiple dendritic segments and traced ROIs (yellow ovals). (b) Example trace of fluorescent change during a 20s trial showing mean activity over 10 trials (red) and S.D. (blue) with odor exposure period (green box). (c) Fluorescent change in all ROIs from one animal averaged in a block during passive ethyl tiglate exposure (between white bars) prior to performing the operant conditioning easy task. (d) Fluorescent change in all RIOs from one animal averaged for the S+ (ethyl tiglate) odorant in a block while performing ≥85% in the easy operant conditioning task. ROI rows correspond to those in (c).

However, comparing this activity between first odor exposure, without reward, and in animals that performed >85% correct in the operant conditioning task (see Fig. 5c,d) showed no obvious change in activity when taking the mean of activities of all 10 trials within a block. Animals learned at different rates, which also complicated finding a correlation between behavior and activity (Fig. 4).

Volcano plotting in Fig. 6 demonstrated an increase in the number of significant GC ROIs becoming more, or less active when comparing between passive exposure (a) and an animal that learned the operant conditioning task (b). Moreover, the percentage of active cells significantly increased when comparing passive exposure and operant conditioning (c).

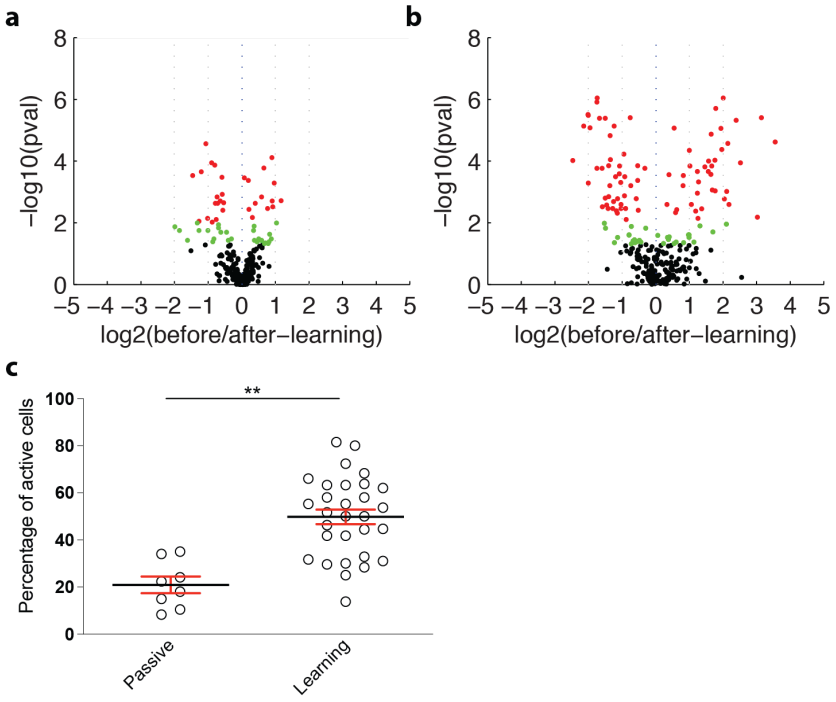


Figure 6: (a) Volcano plot of average fluorescent change (log2) comparing each ROI fluorescent change between ethyl tiglate initial passive odor exposure and last odor exposure (datapoints: black n.s., green P < 0.05, red P < 0.01). (b) Volcano plot of average fluorescent change (log2) comparing each ROI fluorescent change between ethyl tiglate (S+) in operant conditioning task before and after learning (datapoints: black n.s., green P < 0.05, red P < 0.01). (c) Percentage of cells showing significant activity in a passive task and after learning (** p < 0.01).

Unfortunately, this form of analysis is unsuitable for comparing animals and different learning tasks. Due to each animal having a different random spatial distribution of the ROIs, different GCaMP6f labeling efficiency, and differing learning rates, we decided to develop an unbiased, global machine learning approach to find a signature of learning across animals.

3.2 Trial Encoding

We trained an autoencoder (AE) and a variational AE (VAE) neural network to learn a compressed, low-dimensional representation of each time-series of ROI activity. The trained models are used to encode all ROI traces and reduce them into a 4-size vector. The architecture of each network is shown in Fig. 7a,c and two example input time-series are shown in Fig. 7b., in green a non-informative trace and in blue an informative one according to the criteria defined in 2.2.4. The quality of the reconstruction from the 4-dimensional vector is assessed through the mean squared error (MSE). The evolution of the training error for the AE is shown in panel d and the MSE for both models is shown in Fig. 7e. Very similar scores are obtained: 1.30e-2 for the AE and 1.33e-2 for the VAE. Figure 7f–h displays 3 different examples of trace reconstruction by the AE and VAE. One can first note that both models act as noise filters, removing high frequency oscillations. Moreover, both odor (first trace) and reward-evoked (second trace) activities seem well reconstructed, even though individual peaks are not precisely resolved.

Spontaneous activity in the third trace was not reconstructed by either model. The consequence of using a very small latent space is the reduction of the possibility to encode all the diversity of temporal traces. We show below examples of the limitations of representation by these autoencoders.

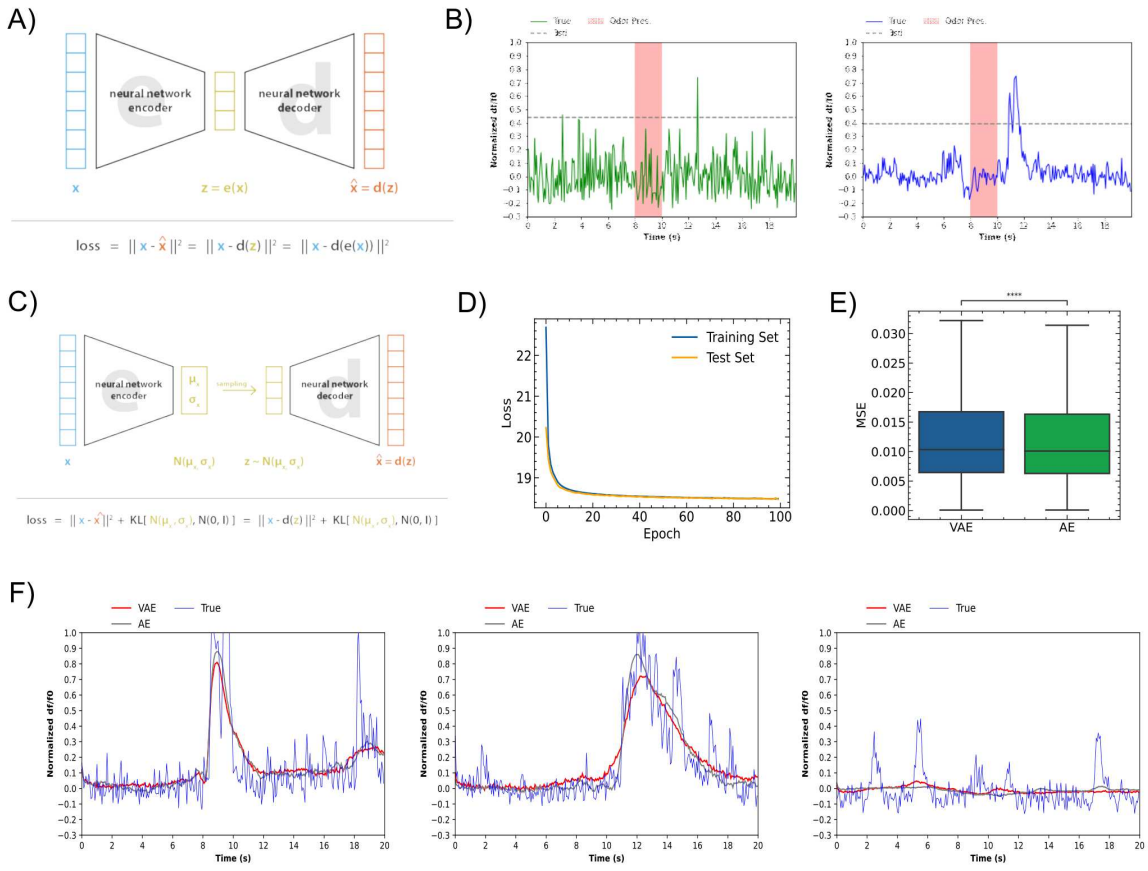


Figure 7: (a) Architecture of the autoencoder (AE). (b) In green a non-informative trace and in blue an informative trace. Criteria: exceeding at least 7 times a threshold of 3 standard deviations (in dotted line). (c) Architecture of the variational autoencoder (VAE). (d) Example of evolution of training and testing loss of the AE. (e) Mean square error of both the AE and the VAE for the entire ROI dataset encoding. (f), (g) and (h) show 3 examples of temporal signal reconstruction from the latent space of the AE and the VAE. Note that some signals are improperly encoded.

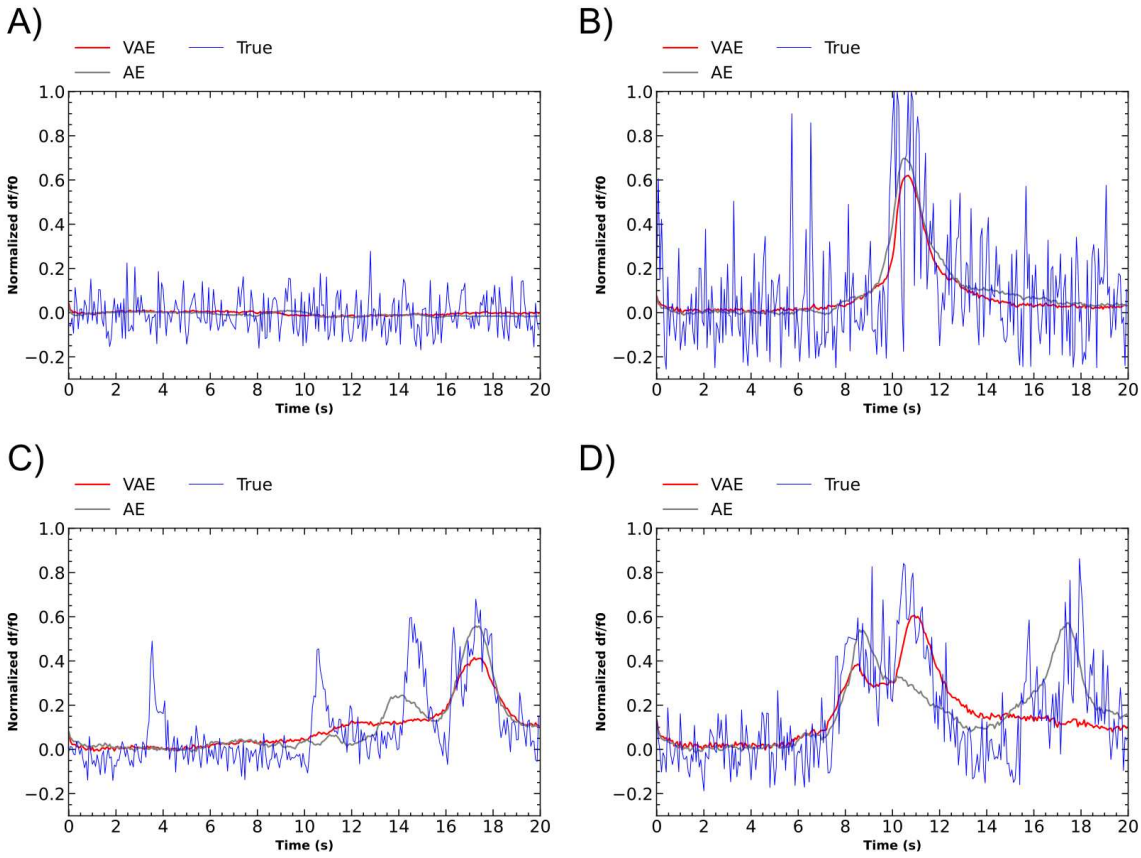


Figure 8: Comparison between raw signal and their reconstitution from the 4D latent vector. In blue the experimental curve, in gray the reconstruction from the AE and in red the reconstruction from the VAE. Temporal evolution of a ROI with no activity in (a) with a low signal to noise ratio in (b) with spontaneous activity in (c) and a more complex response to odor in (d).

3.3 Properties of the latent space

3.3.1 ROI Clustering

The latent space is a 4 dimensional vector space. Each ROI can thus be represented as a distribution of 4-dimensional vectors, in which each realization is a trial. We can exploit the regularity in the latent space of the VAE to define a distance between ROIs. We use the maximum mean discrepancy (MMD)[40–42] as a distance metric. MMD

allows distance calculation between two distributions and requires no knowledge of the underlying probability density function.

We present in Fig. 9a the distance matrix between ROIs for an animal for which the MMD was evaluated across 1207 trials. We performed agglomerative clustering[43] on this matrix; the results are shown in Fig. 9b after rearrangement of the ROIs to correspond to the clustering. While it is challenging to evaluate a precise number of clusters, many ROIs seem relatively close to one another. By their nature, it is expected to find groups of ROIs showing low distances. They are indeed portions of dendrites from the same FoV, belonging thus potentially to the same neurite and exhibiting similar activity. Such ROI clustering is observed for every animal in the VAE latent space, whereas the clusters are less separated in the AE's latent space. Having shown that autoencoders latent space can give structural insights, we will now focus on its functional interpretation.

We focus here on a single ROI for which each trial can be represented in a 4-dimensional space. Figure 9c presents the distribution of trials in the AE's latent space, for a single animal performing an associative learning task. Each trial is represented by a dot and is color-coded by the odor presented. We observe a strict separation in the latent space between S^+ and S^- trials. We underline here that the AE was fitted with the sole purpose of minimizing trace reconstruction error, and thus in an unsupervised manner regarding behavioral outcome.

Stimulus identity is encoded in the latent space for this particular ROI, hinting at different activity patterns upon S^+/S^- presentation. In Fig. 10a, we can observe a similar encoding of reward collection in the AE's latent space for another ROI, though there is more overlap between the S^+/S^- protocols. In Fig. 10b, trials were color-coded by learning state in the VAE's latent space; the increased regularity is visible as data points are more densely clustered together. A separation between learning states is nonetheless observed, notably in dimensions 2 and 3.

Both the AE and VAE latent spaces offer promising graphical representations, in which trials are naturally spatially segregated by behaviors. We will now provide numerical and statistical estimations of these separations and assess inter-animal generalization of these effects.

3.3.2 Predictive analysis

RFCs were trained as described in the *Methods* section on both AE and VAE latent vectors to predict i) the odor presented during a trial, ii) whether the animal was rewarded or not, and iii) the training level of the animal during the trial. Unlike the AE fitted on the whole dataset, the predictions of the RFC are animal- and trial-specific and based on the latent representation of the ROI activities. Each trial corresponds to N calcium traces, encoded as a $4N$ latent vector.

Figure 11a shows the prediction accuracy for 13 animals performing associative learning tasks, regardless of their level of performance. The separation observed in Fig. 9c and Fig. 10a in the AE's latent space is confirmed by high prediction accuracy for odor (87.8% on average) and for reward (84.1% on average) across animals. There is a notable difference of 12.6 % in odor prediction accuracy between easy and difficult tasks.

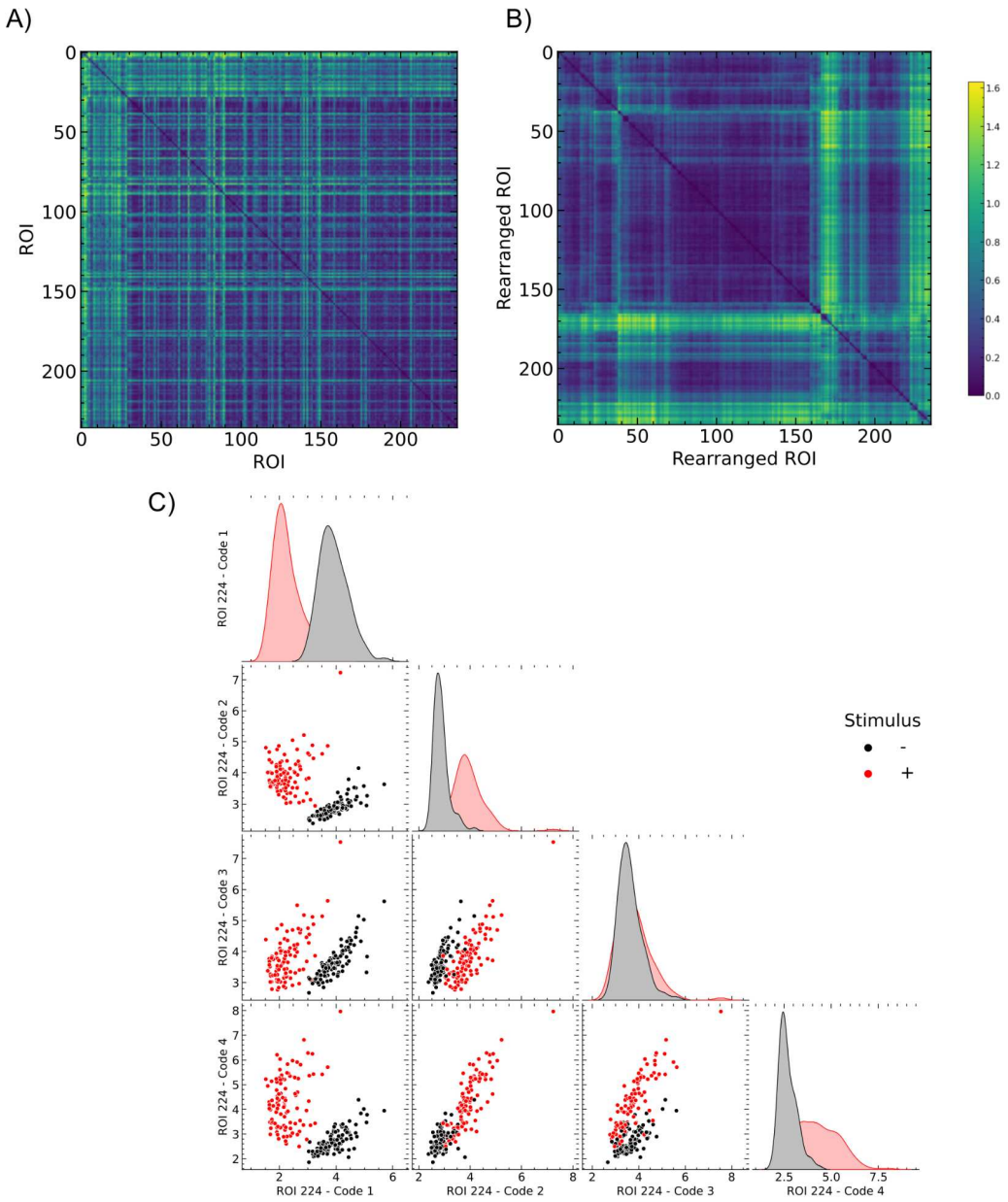


Figure 9: (a) Distance matrix between all ROIs of an animal during all trials. Distance is evaluated as the maximum mean discrepancy from the distribution of values in the learned latent space. (b) Reorganization of the matrix from agglomerative clustering. (c) We show for a single ROI the distribution of trials in the AE latent space, for a single animal performing an associative learning task. Note the natural separation of the trial depending on the S+ and S- stimulus.

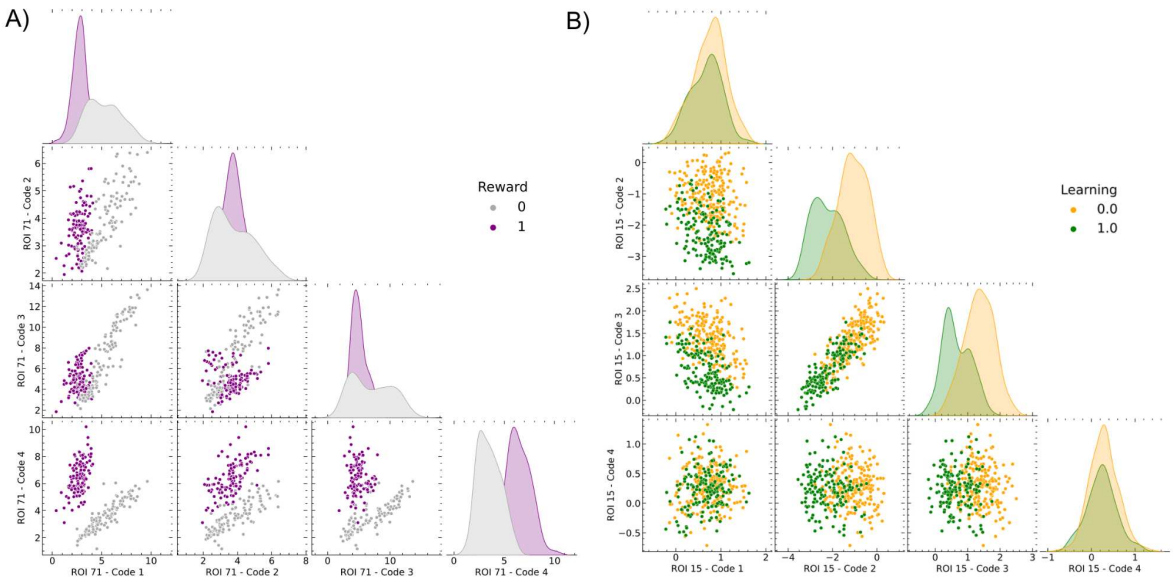


Figure 10: (a) Example of ROI latent representation from the AE color coded by reward delivery. (b) Example of ROI latent representation from VAE color coded by the learning state

Predictors trained on the VAE latent space perform no better than baseline for odor and reward. It is explained in part by the higher regularity of the latent space compared to AE, abolishing inter-trials separation. The opposite effect is seen for learning prediction for which VAE-based accuracy (66%) is higher on average than its AE counterpart (54%). Training on data with shuffled labels yielded an average accuracy corresponding to the random baseline (51%). GLRT performed between the accuracies of the predictions based on the VAE encodings of empirical data and shuffled data rejected the null hypothesis that they were equal. This proves a statistically higher performance compared to a shuffled model, indicating that the neural activity of the labeled GCs are truly predictive of learning and not a spurious effect due to overfitting.

Using RFCs, we showed highly-accurate encoding in AE latent space for odor and reward, and highlighted statistical signatures for learning using a VAE-based compressed

representation. We will try to narrow the scope in time and space, assessing which ROIs and which parts of a trial are most predictive.

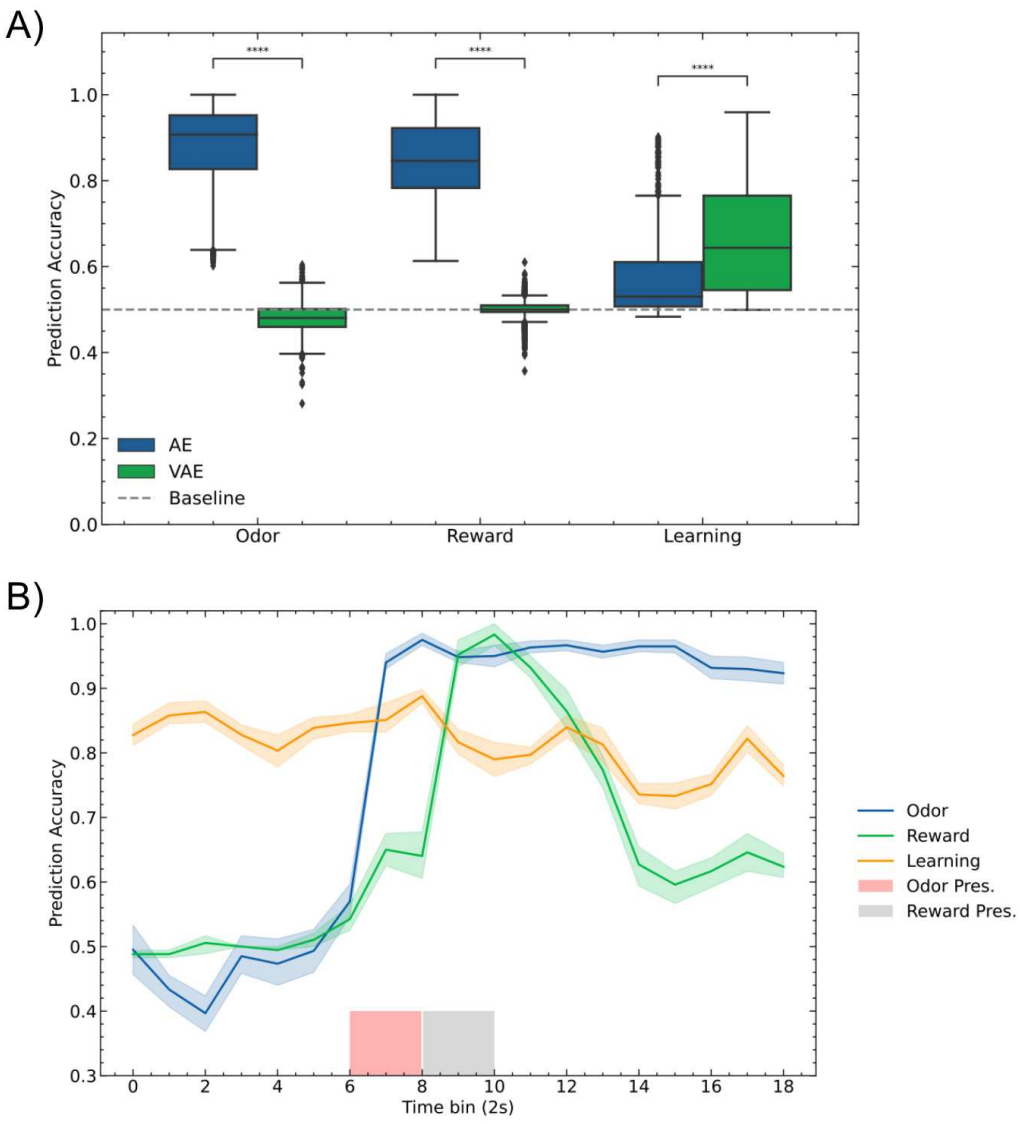


Figure 11: (a) Separation of the information encoded by the AE, which is very performant in encoding odor and reward, and the VAE, which efficiently encodes learning. (b) Prediction accuracy as a function of time for odor, reward and learning prediction. Both odor and reward prediction accuracy matche the temporal nature of the experiments. Learning is a longer process that happens over multiple trials and cannot be associated to a specific time in the trial. Differences between the prediction accuracies based on AE and VAE encodings are all significant at the 0.0001 level (****).

3.3.3 Timing of neural encoding

To define the timing of behavioral encoding in the Ca^{2+} signal, we used a sliding window approach. We encoded bins of 2 seconds corresponding to 30 timeframes, the rest of the signal being set to the baseline 0. This method allows the use of the same autoencoders to cover the whole duration of a trial. We focus here on a single animal performing an associative learning task during 278 trials.

RFCs were trained on each compressed form of successive portions of trials to predict odor, reward and learning. We present in Fig. 11b the corresponding prediction accuracies with their 68% confidence interval. Odor (respectively reward) prediction accuracy remains on the baseline level before odor presentation as expected. Then upon presentation of the odor, the accuracy surges. This localized activity encoding confirms the expected timing of perception. We note a maintained odor prediction accuracy close to maximum after presentation. In comparison, the accuracy of reward prediction decreases faster, hinting for a longer persistence in odor encoding or more limited information encoding in short time windows. Prediction accuracy of the training level of the animal seems less dependent on time, with a steady accuracy around 80% in the given trial. In summary, a precise timing was observed for odor and reward encoding, while learning is not encoded at a specific time during the trials.

3.3.4 Redundancy of neural encoding

We furthermore assessed how information is distributed across ROIs. To do this, we took advantage of the flexible framework offered by the autoencoder-based representation. We could indeed generate different representations of a same trial by including a progressively increasing proportion of ROIs. This procedure enabled the evaluation of the information gain provided by the successive addition of ROIs.

50 subsets of ROIs were randomly selected for each proportion and their corresponding compressed activity was stored. The evolution in prediction accuracy was then assessed by applying the RFC to the corresponding set of latent vectors. Results are available in Fig. 12 over 306 trials during which an animal performs an associative

learning task. Interestingly, maximum accuracy is rapidly reached for all behavioral outcomes, and increases barely with further addition of ROIs. On average only approximately 8% of total ROIs are necessary to achieve 90% of the total odor (respectively reward, learning) prediction accuracy. It shows the high level of redundancy of the information encoded in the different ROIs. This is in agreement with the clustering of ROI activities observed (Section 3.3.1) and may be explained similarly to the clustering: as ROIs are portions of dendrites in a same FoV, an important level of functional redundancy is forecasted. However, we can underline that the confidence intervals have low amplitude, arguing for an even distribution of information across ROIs.

We demonstrated that subsets containing only of the order of 10% randomly selected ROIs are sufficient to almost attain maximal prediction accuracy, we will now investigate whether similar results can be reached using a single ROI. To do so, for each animal we select the ROI whose activity encoding yields the highest prediction accuracy. Distributions of those accuracies are presented in Fig. 12b, showing comparable results as whole population accuracies (Fig. 11a). For instance, the activity of a single portion of dendrite is sufficient to encode odor identity with an average accuracy of 86.5% (averaged across all 13 animals).

We finally investigate the original individual traces of these highlighted ROIs. In Fig. 12c we show normalized fluorescence for one highly predictive ROI for odor over 306 trials. A surge in activity at the 8^{th} second in S+ trials corresponds to the odor presentation, confirming the previously demonstrated encoding. Similarly, Fig. 12d presents the same separation between trials with respect to reward. Here, rewarded trials are associated with a decrease in activity after odor presentation, suggesting inhibition in the corresponding portion of dendrite. AE-based compression allows us to put emphasis on particular ROIs, through assessment of their predictive power. As a comparison, a poorly predictive ROI activity is shown in Fig. 12e. The great similarity between both signals does not enable odor differentiation. Fig. 12f shows the average activity of a ROI that is highly predictive for learning state, their difference being more tenuous than for the highly predictive ROIs for odor and reward though significant on average.

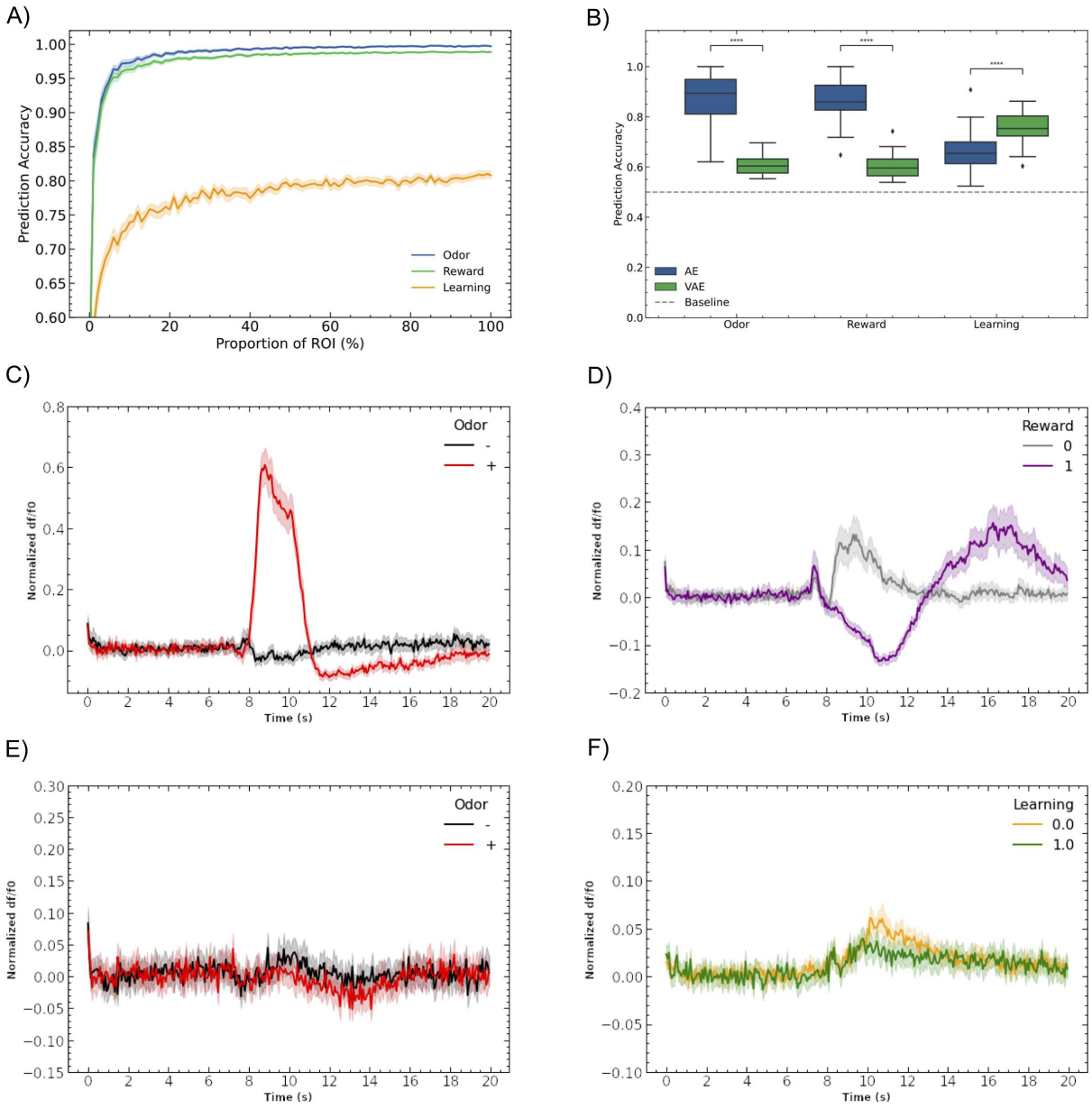


Figure 12: (a) Prediction accuracy as function of the number of ROIs considered. (b) Prediction accuracies based on the AE and VAE encodings for odor, reward and learning when using only the single most predictive ROI. (c–d) Normalized fluorescence traces for a ROI that is highly predictive for odor identity (c) and a ROI that is highly predictive for reward (d). (e) Fluorescence trace for ROI that is poorly predictive for odor identity. (f) Fluorescence trace for ROI that is highly predictive for learning.

4 Discussion

Statistically detectable traces of learning exist in the OB. Despite the sparsity of the marked neurons and the access to only portions of dendrites, the information contained in the individual activity zones (ROIs) is sufficiently redundant to allow us inferring the nature of the odorant, the presence or absence of reward, and the learning level itself. The methods developed in this thesis combine the extensibility offered by an easy-to-train prediction stage, with the expressivity of a common feature space for describing the activity of neuronal populations characterized by areas of interest or activity. Indeed, the latent space is shared by all mice, compiling extensive information, while the prediction stage is plugged-in atop the common latent representations and can thus be easily adapted on a per-mouse or per-task basis.

Interestingly, we observed a difference in properties between AEs and VAEs. AEs performed better at predicting odors and rewards while VAEs performed better at predicting learning. This unusual effect is mainly due to the extreme compression of information in the chosen latent space. While it was chosen to ensure the robustness of the biological results, it has consequences for the properties of the learning procedures. This description differentiation effect is not due to some kind of local maxima in which the AE and VAE would be blocked, most of the learning processes have resulted in this effect. Our approach does not allow a clean quantification of the code redundancy implemented in the image circuit elements. The encoding of information does not suffer from the fact that many ROIs belong to the same dendrites. But the projection into the latent space leads to distortions that hinder reliable identification of the redundancies due to the ROIs belonging to the same dendrites.

Finally, the approach we have introduced is general and not limited to the applications we have explored here. Moreover, it can be scaled up to a large scale, both training and inference can be parallelized as patches can be sent in groups of different sizes independently of each other. Our approach can thus be used in the analysis of circuits where structural differences between animals are important. However, it is not suitable for fine-grained analysis of temporal dynamics in circuits because the embedding takes as input the full temporal series in order to construct the latent space.

References

1. Hogan, M. K., Hamilton, G. F. & Horner, P. J. Neural Stimulation and Molecular Mechanisms of Plasticity and Regeneration: A Review. *Front. Cell. Neurosci.* **14**, (2020).
2. Patch Clamp Techniques for Studying Ionic Channels in Excitable Membranes | Annual Review of Physiology. https://www.annualreviews.org/doi/10.1146/annurev.ph.46.030184.002323?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed.
3. Chen, R., Canales, A. & Anikeeva, P. Neural recording and modulation technologies. *Nat. Rev. Mater.* **2**, 1–16 (2017).
4. Brini, M., Calì, T., Ottolini, D. & Carafoli, E. Neuronal calcium signaling: function and dysfunction. *Cell. Mol. Life Sci.* **71**, 2787–2814 (2014).
5. Tsien, R. Y. New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis, and properties of prototype structures. *Biochemistry* **19**, 2396–2404 (1980).
6. Russell, J. T. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. *Br. J. Pharmacol.* **163**, 1605–1625 (2011).
7. Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373**, 663–664 (1995).
8. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci.* **96**, 11241–11246 (1999).
9. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
10. Sakamoto, M. *et al.* A Flp-dependent G-CaMP9a transgenic mouse for neuronal imaging in vivo. *Cell Rep. Methods* **2**, 100168 (2022).
11. Denk, W., Strickler, J. H. & Webb, W. W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science* **248**, 73–76 (1990).
12. Nimmerjahn, A., Kirchhoff, F. & Helmchen, F. Resting Microglial Cells Are Highly Dynamic Surveillants of Brain Parenchyma in Vivo. *Science* **308**, 1314–1318 (2005).
13. Mizrahi, A. & Katz, L. C. Dendritic stability in the adult olfactory bulb. *Nat. Neurosci.* **6**, 1201–1207 (2003).
14. Mizrahi, A. Dendritic development and plasticity of adult-born neurons in the

mouse olfactory bulb. *Nat. Neurosci.* **10**, 444–452 (2007).

15. Sailor, K. A. *et al.* Persistent Structural Plasticity Optimizes Sensory Information Processing in the Olfactory Bulb. *Neuron* **91**, 384–396 (2016).
16. Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E. & Svoboda, K. Spine growth precedes synapse formation in the adult neocortex in vivo. *Nat. Neurosci.* **9**, 1117–1124 (2006).
17. Holtmaat, A. J. G. D. *et al.* Transient and Persistent Dendritic Spines in the Neocortex In Vivo. *Neuron* **45**, 279–291 (2005).
18. Chu, M. W., Li, W. L. & Komiyama, T. Balancing the Robustness and Efficiency of Odor Representations during Learning. *Neuron* **92**, 174–186 (2016).
19. Andermann, M., Kerlin, A. & Reid, C. Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front. Cell. Neurosci.* **4**, (2010).
20. Reinert, S., Hübener, M., Bonhoeffer, T. & Goltstein, P. M. Mouse prefrontal cortex represents learned rules for categorization. *Nature* **593**, 411–417 (2021).
21. *The Synaptic Organization of the Brain*. (Oxford University Press, 2004). doi:10.1093/acprof:oso/9780195159561.001.1.
22. Nagayama, S., Homma, R. & Imamura, F. Neuronal organization of olfactory bulb circuits. *Front. Neural Circuits* **8**, (2014).
23. Nunes, D. & Kuner, T. Disinhibition of olfactory bulb granule cells accelerates odour discrimination in mice. *Nat. Commun.* **6**, 8950 (2015).
24. Alonso, M. *et al.* Activation of adult-born neurons facilitates learning and memory. *Nat. Neurosci.* **15**, 897–904 (2012).
25. Lepousez, G. *et al.* Olfactory learning promotes input-specific synaptic plasticity in adult-born neurons. *Proc. Natl. Acad. Sci.* **111**, 13984–13989 (2014).
26. Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. (Springer, 2011).
27. Ding, J., Luo, A. F., Hu, L., Wang, D. & Shao, F. Structural basis of the ultrasensitive calcium indicator GCaMP6. *Sci. China Life Sci.* **57**, 269–274 (2014).
28. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
29. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
30. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification And Regression Trees*. (Routledge, 2017). doi:10.1201/9781315139470.

31. Hagerbaumer, J. B. The Gini Concentration Ratio and the Minor Concentration Ratio: A Two Parameter Index of Inequality. *Rev. Econ. Stat.* **59**, 377–379 (1977).
32. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
33. Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science* **313**, 504–507 (2006).
34. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).
35. Kingma, D. P. & Welling, M. *Auto-Encoding Variational Bayes*. (2013).
36. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. *ArXiv14014082 Cs Stat* (2014).
37. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *ArXiv13126114 Cs Stat* (2014).
38. Bishop, C. *Pattern recognition and machine learning*. (Springer, 2006).
39. Chollet, F. & others. Keras. https://github.com/fchollet/keras (2015).
40. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. & Smola, A. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* **13**, 723–773 (2012).
41. Sutherland, D. J. *et al.* Generative Models and Model Criticism via Optimized Maximum Mean Discrepancy. *ArXiv161104488 Cs Stat* (2021).
42. Verdier, H. *et al.* A maximum mean discrepancy approach reveals subtle changes in α-synuclein dynamics. 2022.04.11.487825 (2022) doi:10.1101/2022.04.11.487825.
43. Patel, S., Sihmar, S. & Jatain, A. A study of hierarchical clustering algorithms. in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)* 537–541 (2015).

Collaboration: Hematopoietic stem cell transplantation and modeling

During the PhD, a large-scale project headed by Kurt Sailor within the laboratory Perception and Memory required some modeling to interpret biological findings that were inconsistent with known processes. I, and members of the Decision and Bayesian Computation laboratory modeled some elements of cell proliferation in two distinct brain cell populations. We introduce here the subject of the study, the results of the modeling and the published paper.

1 Introduction

Hematopoietic stem cell transplantation (HSCT), commonly called “bone marrow transplantation” is a clinical treatment for various malignant and non-malignant disorders with annually over 50,000 procedures per year[1]. HSCT has been used to treat a rare genetic disorder, adrenoleukodystrophy, a devastating disease in children that causes rapid demyelination leading to behavioral and speech difficulty, transforming into a catatonic state and, if untreated, is lethal. The driving question is how can a procedure that replaces bone marrow stem cells cure a neurological disease since the brain is normally impermeable to bone-marrow derived cells?

In this study we aimed to determine the mechanism of HSCT on the resident immune cells of the brain, microglia. HSCT treatment requires first depleting the patient’s bone marrow of stem cells using the myeloablative agent busulfan (Fig. 13). This drug covalently cross links DNA residues and is lethal to hematopoietic stem cells. Once the bone marrow niche is depleted, the patient is injected intravenously with hematopoietic stem cells harvested from a matching donor, which re-supplies the recipient’s bone marrow niche with donor stem cells. We suspected busulfan to be responsible for making the brain permissive to peripheral blood cells.

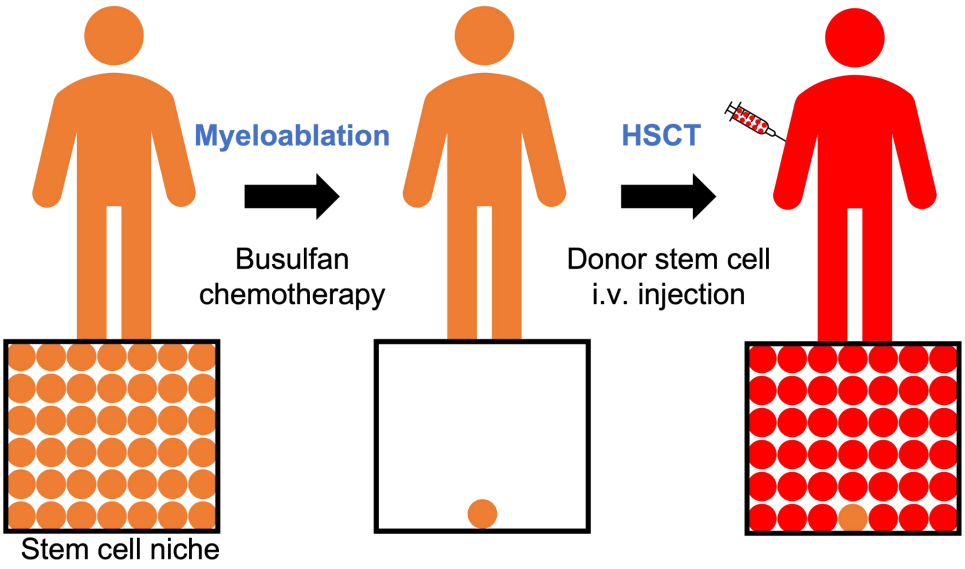


Figure 13: Hematopoietic stem cell transplantation procedure. The bone marrow niche contains blood stem cells (orange circles) and prior to HSCT the stem cell niche needs to be cleared, often using busulfan chemotherapy. Almost all stem cells are depleted and the patient receives an intravenous injection of donor HSCs (red circles) where they home to the niche and become resident stem cells providing the patient with all the bone marrow-derived blood cell types.

2 Methods

2.1 HSCT transplantation and immunohistochemistry

Adult male wild-type or CX3CR1-GFP mice underwent HSCT and at 1 day, 1, 2, 4, 6, 12 and 24 weeks post-HSCT the mice were transcardiac perfused with PFA, the brains dissected out and sectioned. Immunohistochemistry was performed for IBA-1 (macrophage/microglia marker), Ki67 (cell cycle G1-S-G2 marker), and doublecortin (DCX, neuronal progenitor marker). The brain sections were imaged using a confocal microscope and microglia and neural progenitors were quantified using FIJI (ImageJ[2]) software.

2.2 Microglia Ki67 density and age-related decline neurogenesis modeling

Statistical analysis and modeling were performed in Python and goodness-of-fit was assessed with R^2 and root mean square error (RMSE). Since Ki67 is expressed for ~32 hours in microglia[3], the microglia proliferation rate was fitted over time based on the mean of measured data for each time point. Spline interpolation of order 3 was used to ensure the proliferation rate remained positive (R^2 = 0.74 and RMSE = 2.64), with null first-order derivative as constraints at the first and sixth time points (day 1 and week 12). Microglia density (cells/mm²) was fitted over the 5 first time points (until week 6). We considered two separate microglial populations: N_1, the resident microglia, and N_2, the newly generated microglia. A distinct rate model was used to fit each population's evolution, with λ_1 and λ_2 the death rates for N_1 and N_2 respectively. The proliferation rate of population 2, $\mu(t)$, was interpolated from data. For N_1, our model defined the following mass-action rate equation: $\frac{dN_1}{dt}(t) = -\lambda_1 N_1(t)$, leading to $N_1(t) = N_1^0 \exp(-\lambda_1 t)$. Similarly, for N_2 we obtained: $\frac{dN_2(t)}{dt} = \mu(t) - \lambda_2 N_2(t)$. This latter equation was solved iteratively using Euler's method, with a time step (dt) of 1 day following the equation: $N_2(t+dt) = N_2(t)(1-\lambda_2)\,dt + \mu(t)\,dt$. To fit the total microglia density $N(t) = N_1(t) + N_2(t)$, we iteratively optimized λ_1 and λ_2 using the least squares method, and obtained λ_1 = 5.1%, λ_2 = 10.8% and $N_1(0) = 281.8$ (R^2 = 0.84 and RMSE = 27.46).

To fit our control neurogenesis Ki67⁺ and DCX⁺ cell quantification to previously published age-related decline results[3], data were fitted to a distinct rate model with N the population size and λ the corresponding monthly death rate: $\frac{dN}{dt} = -\lambda N(t)$ leading to: $N(t) = N^0 \exp(-\lambda t)$. For Ki67⁺ cells, λ = 39.7% and $N_1(0)$ = 8250 (R^2 = 0.91 and RMSE = 642) and DCX⁺ cells λ = 33.7% and $N_1(0)$ = 28566 (R^2 = 0.89 and RMSE = 2435).

3 Results

3.1 HSCT effect on adult neurogenesis and microglia

Adult neurogenesis is a process of new neuron addition in the adult brain, which in rodents is confined to the hippocampus dentate gyrus and olfactory bulb. Resident stem cells in the subgranule zone of the dentate gyrus and in the subventricular zone of the lateral ventricles continuously divide into neural progenitors that gradually mature and integrate into the circuitry. We assessed the status of these neurogenic zones after HSCT and discovered an almost immediate, complete and permanent loss of all adult neurogenesis within both zones (Fig. 14).

Adult neurogenesis undergoes an age-related decline[4] and it was possible the HSCT effect accelerated this process. We modeled our data to correspond to our control density of cells to show a normal age-related decline. An overlay of the raw Ki67 data was plotted (Fig. 14a, orange open circles) and a clear difference was observed between the modeled decline and the HSCT effect on adult neurogenesis suggesting a very rapid either senescence, or death of resident stem cells. The same dramatic loss was seen in the neural progenitor doublecortin (DCX) positive cells (Fig. 14b, orange open circles). These results show loss of neurogenesis due to cell death, and not via a mechanism accelerating the normal age-related decline.

In the post-HSCT microglia we also modeled the death rate of brain resident microglia based on their density and expression of a proliferation marker (Ki67). Initially there was a ~5% daily death rate in the period immediately after HSCT to 2 weeks post-HSCT (Fig. 14c). With the increase in Ki67 positive microglia at 4-6 weeks post-HSCT, the death rate doubled to ~10%, to allow new microglia addition without changing the overall density of microglia. Once Ki67 ceased at 12 and 24 weeks, we should have observed a near complete collapse in the microglia density at 12 weeks with the ~10% death rate, yet the density remained statistically insignificant as compared to 6 weeks. This model drew into question whether the commonly used proliferation marker Ki67 was faithfully representing cell proliferation and opened our exploration of further

cell cycle markers. This study was published in the March 2022 edition of *Nature Medicine*[5].

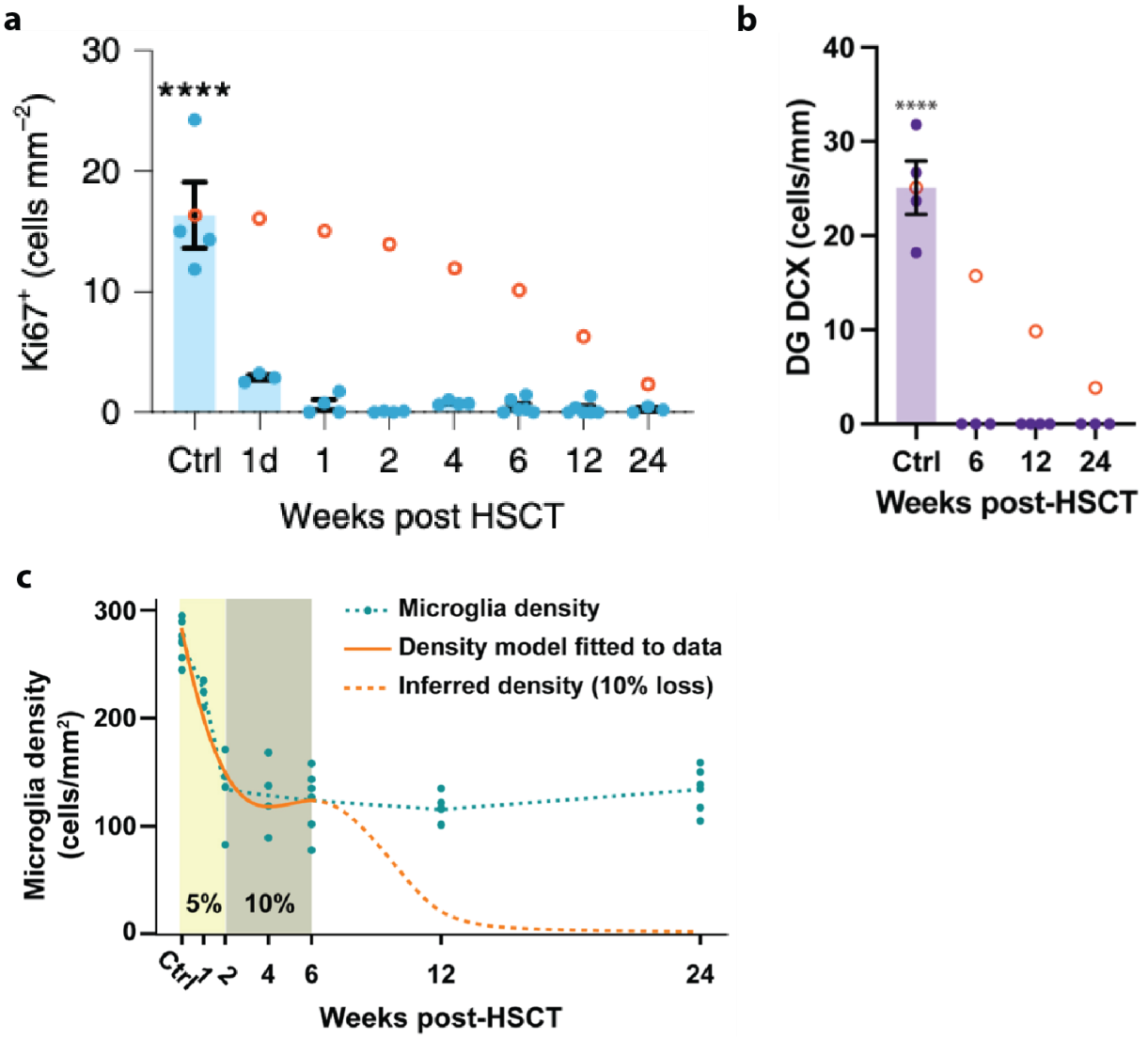


Figure 14: The effect of HSCT on adult hippocampal neurogenesis and microglia proliferation. (a) Plot of hippocampus dentate gyrus Ki67+ cell density (blue filled circles) in control (Ctrl) and at 1 day, 1, 2, 4, 6, 12 and 24 weeks post-HSCT showing significant loss (**** $p < 0.001$ one way ANOVA; mean ± SEM) which was greater than the predicted age-related decline of adult neurogenesis (orange open circles). (b) Plot of doublecortin density in hippocampal dentate gyrus region (purple filled circles) in control (Ctrl) and at 6, 12 and 24 weeks post-HSCT showing a significant loss (**** $p < 0.001$ one way ANOVA; mean ± SEM) which was greater than the modeled age-related decline of adult neurogenesis (orange open circles). (c) Plot of experimentally measured microglia density (green filled circles and dotted line) at control (Ctrl), 1, 2, 4, 6, 12, 24 weeks post-HSCT with modeling based on Ki67 presumed proliferation (orange line) and the effect of lost proliferation showing the inferred density (orange dashed line). Sub-figures from Sailor et al., Nat Med 28, 517-527 (2022).

References

1. Aljurf, M. *et al.* “Worldwide Network for Blood & Marrow Transplantation (WBMT) special article, challenges facing emerging alternate donor registries”. *Bone Marrow Transplant.* **54**, 1179–1188 (2019).
2. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
3. Askew, K. *et al.* Coupled Proliferation and Apoptosis Maintain the Rapid Turnover of Microglia in the Adult Brain. *Cell Rep.* **18**, 391–405 (2017).
4. Ben Abdallah, N. M.-B., Slomianka, L., Vyssotski, A. L. & Lipp, H.-P. Early age-related changes in adult hippocampal neurogenesis in C57 mice. *Neurobiol. Aging* **31**, 151–161 (2010).
5. Sailor, K. A. *et al.* Hematopoietic stem cell transplantation chemotherapy causes microglia senescence and peripheral macrophage engraftment in the brain. *Nat. Med.* 1–11 (2022) doi:10.1038/s41591-022-01691-9.

nature medicine

ARTICLES

https://doi.org/10.1038/s41591-022-01691-9

Check for updates

Hematopoietic stem cell transplantation chemotherapy causes microglia senescence and peripheral macrophage engraftment in the brain

Kurt A. Sailor[1]✉, George Agoranos[1], Sergio López-Manzaneda[2,3,8], Satoru Tada[2,3,8], Beatrix Gillet-Legrand[2,3,8], Corentin Guerinot[1,4,5], Jean-Baptiste Masson[4], Christian L. Vestergaard[4], Melissa Bonner[6], Khatuna Gagnidze[6], Gabor Veres[6], Pierre-Marie Lledo[1,8]✉ and Nathalie Cartier[2,3,7,8]✉

Hematopoietic stem cell transplantation (HSCT) is a therapy used for multiple malignant and nonmalignant diseases, with chemotherapy used for pretransplantation myeloablation. The post-HSCT brain contains peripheral engrafted parenchymal macrophages, despite their absence in the normal brain, with the engraftment mechanism still undefined. Here we show that HSCT chemotherapy broadly disrupts mouse brain regenerative populations, including a permanent loss of adult neurogenesis. Microglial density was halved, causing microglial process expansion, coinciding with indicators of broad senescence. Although microglia expressed cell proliferation markers, they underwent cell cycle arrest in S phase with a majority expressing the senescence and antiapoptotic marker p21. In vivo single-cell tracking of microglia after recovery from chemical depletion showed loss of their regenerative capacity, subsequently replaced with donor macrophages. We propose that HSCT chemotherapy causes microglial senescence with a gradual decrease to a critical microglial density, providing a permissive niche for peripheral macrophage engraftment of the brain.

Since 2000, over 400,000 HSCTs have been performed in the United States as a therapy for multiple malignant and nonmalignant diseases[1], with chemotherapy agents commonly used for myeloablation[2]. Chemotherapy kills recipient bone marrow cells before transplantation to provide donor stem cells with a permissive niche to engraft the recipient's bone marrow. HSCT is also a therapy used for leukodystrophies[3–5] but, since peripheral cells do not breach the blood–brain barrier (BBB) in the normal brain, it is unclear how HSCT could be therapeutic for these white matter disorders. Recent studies in both mice and humans have demonstrated that donor cells gradually accumulate in the post-HSCT brain as parenchymal macrophages, suggesting they provide a supportive role in central nervous system therapies[1,6].

It has been proposed that brain conditioning with either irradiation or chemical myeloablative procedures is required for peripheral cell brain engraftment. HSCT, in combination with total body irradiation while shielding the head, or using the BBB-impermeable myeloablative drug treosulfan, both failed to engraft donor cells into the brain[7–9]. When irradiation was conversely focused on the head, without myeloablation, this was sufficient to allow donor cell brain engraftment[10]. These and other studies suggest inflammatory signaling, or BBB changes due to irradiation or chemotherapy-induced brain conditioning, as the mechanism engrafting peripheral cells into the brain.

Recent studies,using genetic methods to deplete microglia, were able to engraft peripheral cells without performing irradiation or chemical myeloablation. Genetic ablation of microglia under the control of the CX3C chemokine receptor 1 (CX3CR1) promoter, followed with either intravenous monocyte injections, HSCT with head-shielded total body irradiation or parabiosis, caused donor cells to engraft as brain parenchymal macrophages[11,12]. These studies support a brain microglia-depletion, permissive niche mechanism, yet it remains unclear whether this occurs with chemotherapy-based HSCT.

Here, we focused on a clinical dose of busulfan chemotherapy myeloablation to determine how this HSCT regimen affects microglia and other brain regenerative populations, to determine the mechanism of peripheral cell brain engraftment.

Results

HSCT donor cells become brain-resident macrophages. To first determine the fate of donor cells in the post-HSCT brain, recipient wild-type C56BL/6j mice, or mice with expression of green fluorescent protein (GFP) in monocytes/macrophages/microglia (CX3CR1-GFP), were myeloablated with busulfan for 4 days (25 mg kg^{-1} day^{-1} intraperitoneally (i.p.; 100 mg kg^{-1} total dose)), a dose replicating clinical levels[13], and transplanted with tdTomato (tdTom)-expressing, lineage-negative donor cells 1 day after the final busulfan dose (Fig. 1a). For the in vivo imaging cohort, a chronic cranial window was created over the somatosensory/motor cortex (dorsal cortex) of CX3CR1-GFP mice 4 weeks before

[1]Institut Pasteur, Université de Paris, CNRS UMR 3571, Perception and Memory Unit, Paris, France. [2]INSERM UMR1169, Université Paris-Sud, Université Paris-Saclay, Orsay, France. [3]Thérapie Cellulaire et Génique des Maladies Neurologiques de l'Enfant et de l'Adulte, Institut du Cerveau et de la Moelle Épinière, Hôpital Pitié-Salpêtrière, Paris, France. [4]Decision and Bayesian Computation, USR 3756 & Neuroscience Department CNRS UMR 3751, Institut Pasteur, Université de Paris, CNRS, Paris, France. [5]Sorbonne Université, Collège doctoral ED3C, Paris, France. [6]bluebird bio, Inc., Cambridge, MA, USA. [7]Present address: Asklepios Biopharmaceutical, Inc., Institut du Cerveau (ICM), Paris, France. [8]These authors contributed equally: Sergio López-Manzaneda, Satoru Tada, Beatrix Gillet-Legrand, Pierre-Marie Lledo, Nathalie Cartier. ✉e-mail: ksailor@pasteur.fr; pierre-marie.lledo@pasteur.fr; ncartier@askbio.com

transplantation, and host microglia/macrophages (GFP$^+$) and donor-derived cells (tdTom$^+$) were imaged over multiple time points (Fig. 1a,b). Immunohistochemistry analysis was performed on the dorsal cortex for Iba-1, a microglia/monocyte/macrophage marker, and TMEM119, a marker specific to brain microglia, 24 weeks post HSCT (24 w; Fig. 1c). TdTom$^+$ donor-derived cells first appeared in the brain parenchyma at 6 weeks post HSCT and increased significantly over the subsequent 18 weeks (Fig. 1d). All tdTom$^+$ donor cells were TMEM119 negative and most (~90%) were Iba-1 positive, as observed previously[6,11], with the remaining tdTom$^+$ cells most probably myeloid-derived cells trapped in the vasculature (Extended Data Fig. 1a)[6,11]. There were regional differences in donor engraftment efficiency, with the olfactory bulb having the highest and the striatum the lowest donor cell density (Extended Data Fig. 1b).

In the first week post HSCT, in vivo two-photon imaging detected tdTom$^+$ dynamic meningeal phagocytes in the arachnoid space (Supplementary Video 1). The first tdTom$^+$ parenchyma donor cells were found in vivo at 8 weeks post HSCT (Supplementary Video 2). The size of donor macrophages did not significantly change over the observation period, and they maintained their radial resting morphology (Extended Data Fig. 1c,d). Single-cell three-dimensional (3D) tracking of 43 donor cells in the same brain volume showed their accumulation (Fig. 1e,f), with 40 cells persisting (Fig. 1f lower plot, black circles) and three cells either dying or migrating out of the imaging volume (Fig. 1f lower plot, red circles). Thus, with their continued presence, we conclude that most donor cells became resident.

Imaging every 10 min showed the macrophage protrusions to be highly dynamic (Fig. 1g and Supplementary Video 3), persistently surveilling their environment with ~30% change (Fig. 1h), higher than control host microglia (~18% change). These in vivo results show macrophages becoming resident, with a resting morphology and surveilling their environment, possibly fulfilling a similar role to microglia despite having a significantly different transcriptome profile[6,11,14,15].

Post-HSCT microglia partially deplete and enlarge. We next tracked host microglial changes in response to busulfan chemotherapy. Using the same experimental outline (Fig. 1a) host microglia were characterized, adding time points at 1 day and 1 week post HSCT for immunohistochemistry (tdTom$^-$/Iba-1$^+$/TMEM119$^+$; Fig. 1c). Host microglial density insignificantly declined after the final busulfan administration at 1 day to 1 week (Fig. 2a). By 2 weeks post HSCT, a significant (~50%) loss of host microglial density was sustained up to 24 weeks. These results were confirmed in vivo with two-photon imaging of CX3CR1^{-}GFP HSCT mice (Extended Data Fig. 2a,b). Host-derived macrophages were also found in the brain parenchyma after HSCT due to incomplete donor chimerism, but this population was >100 times rarer than donor macrophages (Extended Data Fig. 2c,d), demonstrating that peripheral engraftment is not exclusive to donor cells.

Convex hull tracing of individual host microglial processes was performed (Fig. 2b, dashed lines) and, from 4 to 24 weeks post HSCT, their individual coverage area doubled (Fig. 2c). Complementary tiling between host microglia and donor macrophages was observed in vivo, with borders clearly delineated (Fig. 2d). There was an inverse correlation between host microglial density and cell area (Fig. 2e), indicating that microglia enlarge in response to a decrease in density. At 24 weeks post HSCT, higher donor macrophage density also correlated with decreased host microglial density, as expected with complementary tiling (Extended Data Fig. 2e). The higher presence of donor macrophages in the brain also correlated with increased host microglial area at 24 weeks (Fig. 2f).

Donor macrophage area remained the same between 12 and 24 weeks (Extended Data Fig. 2f), comparable to that of host-derived macrophages (Extended Data Fig. 2g). Since all individual donor cells were of the same area, while never having been exposed to busulfan, they probably proliferated to maintain brain tiling. In contrast, host microglial density decline, coupled with increased process coverage, could have been a result of microglial proliferation deficits due to busulfan[16,17].

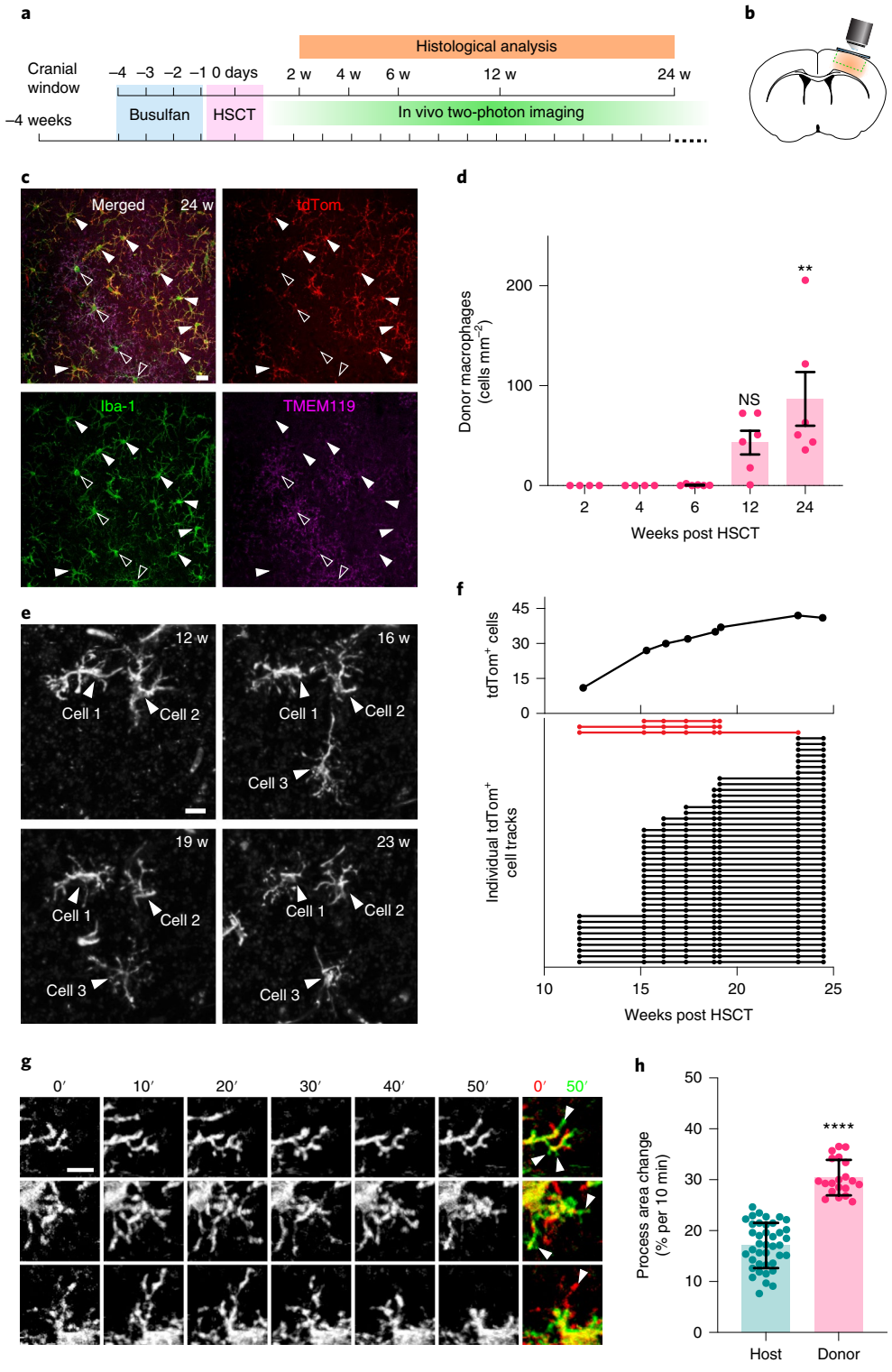
HSCT chemotherapy disrupts cell proliferation in the brain. Since busulfan chemotherapy causes crosslinking of DNA and has been shown to induce cell cycle arrest[18], we first looked at the effect of HSCT on adult neurogenesis, a persistently mitotic cell population. In the hippocampal dentate gyrus, mitotic stem/progenitor cells—as demonstrated by the proliferation marker Ki67 in the subgranular zone—were observed in control mice (Fig. 3a), but at 1 day post busulfan there was a ~80% decrease. This declined to near-complete loss of proliferation, sustained to 24 weeks post HSCT (Fig. 3a,b), diminishing more rapidly than expected with normal age-related decline[19] (Fig. 3b, orange circles). The same loss occurred in subventricular zone (SVZ) Ki67$^+$ stem/progenitor cells (Extended Data Fig. 3a,b). Labeling for doublecortin, an immature neuron/progenitor marker, showed ablation in regions of adult-born neuronal integration: the hippocampus and olfactory bulb (Extended Data Fig. 3c–g). Thus we demonstrate a chemotherapy agent having a permanent effect on extinguishing all adult neurogenesis. Interestingly, oligodendrocyte progenitor cells (NG2), another regenerating population, suffered near-complete loss of cell density at 2 and 4 weeks post HSCT but had begun to recover by 6 weeks post HSCT (Extended Data Fig. 4), demonstrating diverse effects of busulfan chemotherapy on different regenerating cell populations of the brain.

We next investigated the proliferation of donor macrophages and host microglia post HSCT. Donor macrophages showed an insignificant increase in Ki67 expression at 2–24 weeks post HSCT (Extended Data Fig. 5a). Host microglia showed a transient,

Fig. 1 | HSCT causes the accumulation of brain parenchymal donor macrophages that become resident and surveillant. a, Experimental timeline for myeloablation with busulfan, HSCT of tdTom$^+$/Lin$^-$ cells with histological and in vivo analysis at indicated time points. **b**, Coronal brain section illustration of approximate area for histological analysis (orange) and in vivo two-photon imaging through a chronic cover glass cranial window (green dashed line). **c**, Confocal images of immunohistochemistry at 24 weeks for Iba-1 and TMEM119 showing donor macrophages (tdTom$^+$/Iba-1$^+$/TMEM119$^-$, closed arrowheads) and host microglia (tdTom$^-$/Iba-1$^+$/TMEM119$^+$, open arrowheads). Scale bar, 20 μm. **d**, Plot of donor macrophage density in the dorsal cortex at 2, 4, 6, 12 and 24 weeks post HSCT (mean ± s.e.m.; one-way two-sided ANOVA with Tukey's multiple comparisons test; $^{**}P = 0.0078$, 0.0078 and 0.0029 for 2, 4 and 6 weeks, respectively versus 24 weeks; $n = 4$ mice for 2 and 4 weeks, $n = 6$ mice for 6, 12 and 24 weeks). **e**, In vivo two-photon images of the same donor macrophages at 12, 16, 19 and 23 weeks post HSCT. Scale bar, 20 μm. **f**, Plot of accumulation of donor cells (top) and individual donor cell tracking over 25 weeks (bottom) in one mouse post HSCT with cells that persisted to the final imaging session (black circles/lines) and those that died or migrated out of imaging volume (red circles/lines). **g**, Three in vivo donor resident macrophages in brain parenchyma imaged every 10 min for 60 min, showing the dynamics of their processes. Last panel is an overlay of the first (0′, red) and last timepoints (50′, green) showing net dynamic changes. Scale bar, 10 μm. **h**, Average percentage change in process area during 10 min imaging interval between host microglia and donor macrophages, with significantly greater area change in donor macrophages (mean ± s.d.; unpaired two-sided *t*-test, $^{****}P < 0.0001$; $n = 40$ and 20 cells for host microglia and donor macrophages, respectively). NS, not significant.

significant increase in Ki67 expression at 1 week post HSCT and a significant, gradual increase at 4 and 6 weeks post HSCT (Fig. 3c,d). Ki67 expression had dropped to control levels by 12 and 24 weeks, coinciding with the period of significant donor cell engraftment (Fig. 1d). These results were perplexing since no associated increase in host microglial density (Fig. 2a) concurred with the increased Ki67 expression period.

Using a rate model based on host microglial density (Extended Data Fig. 5b), the initial daily loss rate of microglia was ~5% during the first 2 weeks. Since density was maintained after 2 weeks, the host daily microglial death rate would have needed to increase to ~10% to compensate for increased Ki67 expression, assumed to indicate cell proliferation, thereby forming an almost perfect equilibrium between loss and proliferation to maintain microglial density (Extended Data Fig. 5b). Moreover, the modeled 10% death rate at 6 weeks would infer almost complete depletion of host microglia by 12 weeks with loss of Ki67 expression (Extended Data Fig. 5b, inferred density), which was not experimentally observed. We therefore investigated whether Ki67 expression in host microglia post HSCT truly indicated cell proliferation, since



DNA crosslinking renders cells incapable of DNA synthesis and thus causing cell cycle arrest while expressing Ki67 (ref. [20]).

We explored the cell cycle with multiple markers (Fig. 3e) to determine at which stage microglia undergo arrest, using Ki67 and MCM-2 for G1 through G2, 5-ethynyl-2'-deoxyuridne (EdU) for S phase, proliferating cell nuclear antigen (PCNA) for S phase through G2, phospho-gamma-H2A.X (pγH2A.X) for DNA damage in S phase and phospho-histone H3 (pHH3) as a late G2 marker before mitosis[21]. At 6 weeks post HSCT, the peak of Ki67 expression, mice were injected with the DNA analog EdU 2 days before sacrifice and the ratio of EdU$^+$/Ki67$^+$ cells was measured to determine whether Ki67 expression indicated DNA synthesis[22]. In control mice, the ratio was measured in the SVZ (Fig. 3f) since the cortex was almost completely devoid of Ki67$^+$cells (Fig. 3d, ctrl) and an approximately fivefold greater number of EdU$^+$ cells was detected compared to Ki67$^+$cells (Fig. 3g). In mouse dorsal cortex 6 weeks post HSCT this ratio was the opposite, with around sixfold more Ki67$^+$cells than EdU$^+$ cells (Fig. 3f,g), demonstrating near-complete absence of EdU incorporation into host microglial DNA. Immunostaining for the DNA damage marker pγH2A.X showed the near absence of Ki67 colabeling in control, but it was significantly coexpressed at 6 weeks post HSCT in the dorsal cortex of host microglia (Fig. 3h,i). Both results strongly suggested cell cycle arrest[20].

Ki67$^+$ host microglia at 6 weeks post HSCT were almost completely coexpressed with MCM-2 (Fig. 3j), which was insignificant in control SVZ Ki67$^+$ cells (Fig. 3k), supporting the finding that cells had entered G1. Less than half of 6-week post-HSCT microglia coexpressed Ki67 with PCNA (Fig. 3l), as opposed to near-complete coexpression in control SVZ Ki67$^+$ cells (Fig. 3m), suggesting induction of S phase and possible arrest. Finally, to determine whether there was a transition to G2, Ki67$^+$ and pHH3 coexpression was evaluated. Ki67$^+$ control cells showed ~12% coexpression with pHH3 (Fig. 3n,o) whereas no Ki67$^+$ host microglia at 6 weeks post HSCT expressed pHH3 (Fig. 3o). Taken together, these results show microglial cell cycle arrest at S phase, probably due to busulfan-mediated DNA crosslinking[18].

HSCT chemotherapy causes senescence in the brain. We next stained for senescence-associated beta-galactosidase (SA-β-gal), whose expression was prevalent at 2 and 4 weeks post HSCT across the dorsal cortex as compared to control, but decreased to control levels between 6 and 24 weeks (Fig. 4a,b). Since SA-β-gal is not an absolute marker of senescence[23], we also stained for the nuclear membrane protein lamin B1, which is lost when cells undergo senescence[24]. We saw a significant loss at 2 weeks post HSCT, sustained up to 12 weeks across the dorsal cortex but insignificant at 24 weeks (Fig. 4c,d).

We next stained for the cyclin-dependent kinase inhibitor p21 (Fig. 4e), an indicator of cell cycle arrest and senescence that can result from DNA damage[25]. Microglial nuclear staining for p21 was observed at 6 weeks post HSCT (Fig. 4e) yet was absent in control microglia (Fig. 4f). At 1, 4 and 6 weeks post HSCT, Iba-1$^+$ host microglia showed significantly increased expression of p21 (Fig. 4f). At these time points, the percentage of Iba-1$^+$ combined expression with p21 and/or Ki67 was not significantly different (Fig. 4g), with ~65% of host microglia being p21$^+$/Ki67$^+$, ~27% p21$^-$/Ki67$^+$ and ~7% p21$^+$/Ki67$^-$. Taken together, these results suggests that host microglia commence cell division but undergo cell cycle arrest at S phase, activate p21 and become senescent. We therefore decided to test in vivo the extent of post-HSCT microglial senescence.

Post-HSCT microglia lose all regenerative capacity. To test the regenerative capacity of microglia post HSCT, the CSF1R antagonist PLX3397 was employed to deplete microglia and observe their ability to recover. PLX3397 was previously shown to cause near-complete loss of microglia within 4 weeks of administration and, after drug withdrawal, microglia recovered within 1 week, being exclusively repopulated with microglia with no peripheral cell brain engraftment[26]. CX3CR1-GFP mice with a cranial window were utilized for in vivo two-photon imaging with a control group receiving neither busulfan nor donor cells, and a HSCT group receiving busulfan and transplanted with tdTom$^+$ lineage$^-$ donor cells (Fig. 5a). Both groups were given a PLX3397-containing diet for 4 weeks to deplete microglia, followed by drug withdrawal (normal diet) for recovery. Throughout the experiment the same brain volume for each mouse was two-photon imaged over multiple time points (Fig. 5a,b).

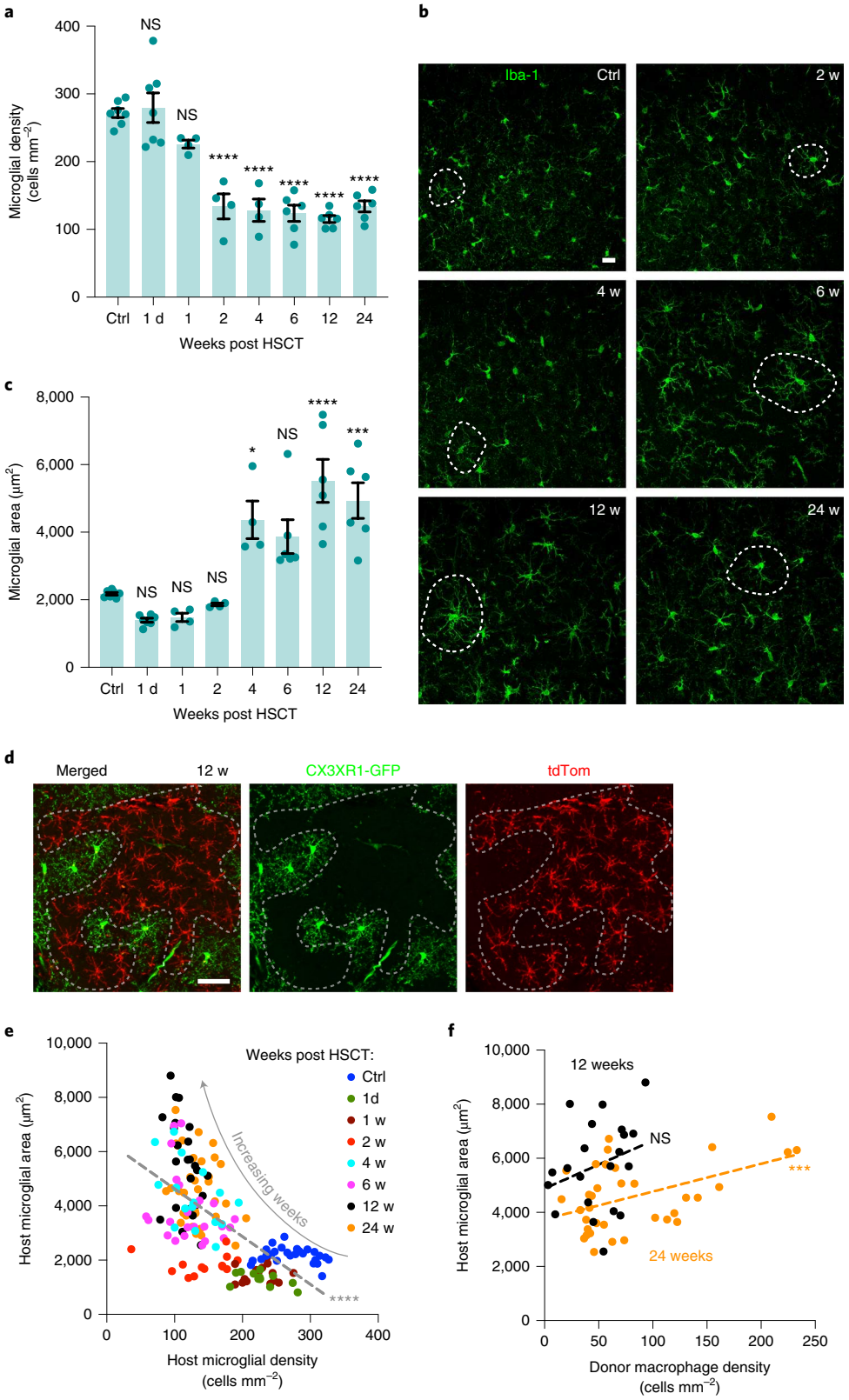
Mice were imaged 4 weeks after cranial window surgery (−7 w) for baseline microglial density and again 3 weeks later (−4 w), immediately before PLX3397 administration. Control microglial density was sustained, but HSCT mice showed a significant decrease due to busulfan treatment (−4 w; Fig. 5c), as in our previous results (Extended Data Fig. 2b). After 4 weeks of PLX3397 treatment, both groups showed significant microglial density loss (0 days; Fig. 5c). The control group recovered within 4 days after PLX3397 withdrawal due to proliferation of host microglia (Fig. 5c), as shown previously[27]. With HSCT there was no recovery of host microglia, even up to 9 days after PLX3397 withdrawal. Moreover, 3D single-cell tracking of all microglia within individual brain volumes demonstrated no cell divisions out of 481 tracked cells in six mice (Extended Data Fig. 6).

At 5 days after PLX3397 recovery in the HSCT group, donor macrophages were first observed in the brain with varying rates of cell engraftment between individual mice (Fig. 5d). Of interest, mouse 2 had the lowest microglial depletion after PLX3397 treatment (microglial density of ~200 cells mm^{-2}; Fig. 5d) and no donor cells were observed engrafting within the imaging window, even at 19 days post PLX where, in contrast, mouse 1 (microglial density of ~50 cells mm^{-2}; Fig. 5d) showed almost complete donor engraftment (Extended Data Fig. 7a). This suggests that a minimal host microglial density threshold is required for successful donor engraftment. Despite this greater host microglial density in mouse 2, none of

Fig. 2 | HSCT busulfan chemotherapy causes loss of host microglia cells and increase in individual microglial process area correlating with donor macrophage density. a, Host microglial density at control (ctrl), 1 day (1d) and 1, 2, 4, 6, 12 and 24 weeks post HSCT (mean ± s.e.m.; one-way, two-sided ANOVA with Tukey's multiple comparisons test; ****$P < 0.0001$ versus ctrl; $n = 7, 7, 4, 4, 4, 6, 6$ and 6 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively). **b**, Host microglia (Iba-1) at ctrl and 2, 4, 6, 12 and 24 weeks post HSCT showing examples of convex hull tracings (dashed lines). Scale bar, 20 μm. **c**, Host microglial process convex hull area at ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks post HSCT (mean ± s.e.m.; one-way, two-sided ANOVA with Tukey's multiple comparisons test; ****$P < 0.0001$, ***$P < 0.001$, *$P < 0.05$; $P = 0.0147$, $P < 0.0001$ and $P = 0.0002$ for 4, 12 and 24 weeks, respectively, versus ctrl; $n = 7, 7, 4, 4, 4, 6, 6$ and 6 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively). **d**, CX3CR1–GFP at 12 weeks post HSCT demonstrating complementary tiling between host (green) and donor (red) cells. Scale bar, 70 μm. **e**, Host microglial density versus host microglial convex hull at ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks post HSCT. Mean of individual brain sections; simple linear regression (dashed line), significance from slope = 0, $P < 0.0001$, $F = 112.6$, $R^2 = 0.4057$; $n = 27, 12, 16, 13, 16, 24, 24$ and 35 brain sections for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively. **f**, Donor macrophage density versus host microglial convex hull at 12 and 24 weeks post HSCT. Mean of individual brain sections; simple linear regression (dashed lines) significance from slope = 0, $P = 0.2165$, $F = 1.635$, $R^2 = 0.07922$ and $P = 0.0042$, $F = 9.448$, $R^2 = 0.2226$ for 12 and 24 weeks, respectively; $n = 24$ and 35 brain sections for 12 and 24 weeks, respectively).

its 233 tracked microglia underwent cell division (Extended Data Fig. 6, mouse 2).

Time-lapse 10-min-interval in vivo imaging showed a period of massive proliferation of donor macrophages within the parenchyma at 6 days after PLX3397 withdrawal in mice with donor engrafted cells (Fig. 5e and Supplementary Video 4). Cells migrated, condensed to a mitotic sphere, developed a cleavage plane, extended their processes and separated a daughter cell (Fig. 5e). There were ~33 proliferating cells mm^{-3} per 10 min (Fig. 5f), peaking to ~115 cells mm^{-3} per 10 min depending on the time period imaged for each mouse, with 7 days post PLX showing the highest level of proliferation (Extended Data Fig. 7b). Of importance, no host



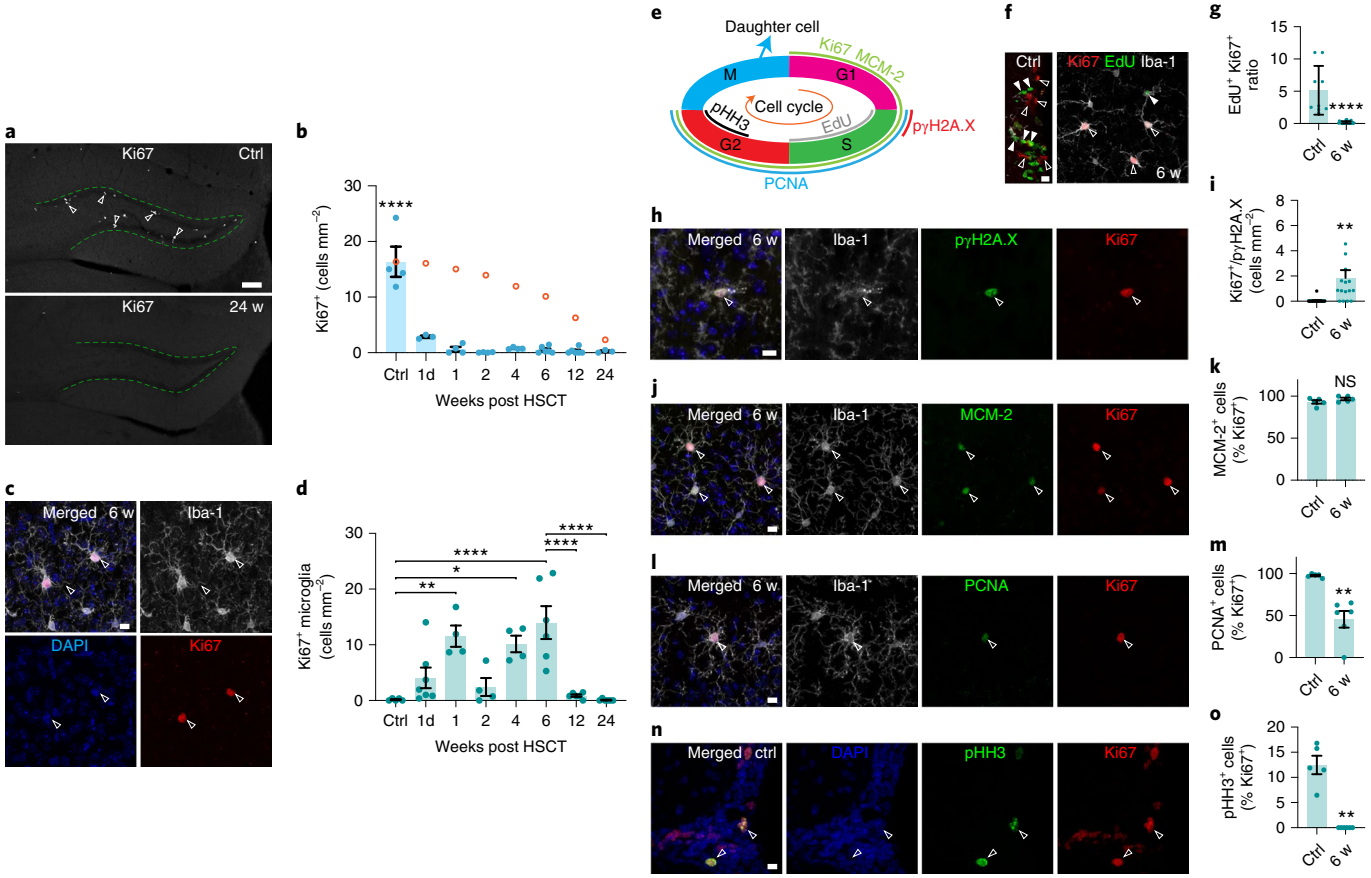


Fig. 3 | HSCT causes loss of adult neurogenesis and host microglial cell cycle arrest. a, Hippocampal dentate gyrus (dashed line) Ki67$^+$ cells (open arrowhead). Scale bar, 100 µm. **b**, Dentate gyrus subgranular zone Ki67$^+$ density with modeled age-related Ki67 decline (orange cicles) (mean ± s.e.m.; one-way, two-sided ANOVA with Tukey's multiple comparisons test; ****$P < 0.0001$ ctrl versus all HSCT time points; $n = 4$, 3, 4, 4, 4, 6, 6 and 3 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively). **c**, Ki67$^+$ (red) microglia (white, open arrowheads). Scale bar, 10 µm. **d**, Plot of Ki67$^+$ microglia at ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks post HSCT (mean ± s.e.m.; one-way, two-sided ANOVA with Tukey's multiple comparisons test; ****$P < 0.0001$, **$P < 0.01$, *$P < 0.05$; $P = 0.0023$, $P = 0.0100$ and $P < 0.0001$ for 1, 4 and 6 weeks, respectively, versus ctrl, $P < 0.0001$ for 12 and 24 weeks versus 6 weeks; $n = 7$, 7, 4, 4, 4, 6, 6 and 3 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively). **e**, Cell cycle markers used for growth 1 (G1), S phase (S), growth 2 (G2) and mitosis (M). **f**, Ki67$^+$ (red, open arrowheads) and EdU$^+$ (green, closed arrowheads) cells. Scale bar, 10 µm. **g**, Ratio of EdU$^+$/Ki67$^+$ cells (mean ± s.d.; two-sided Kolmogorov–Smirnov test, ****$P < 0.0001$; $n = 9$ and 25 brain sections for control and HSCT, respectively). **h**, Ki67$^+$ (red, open arrowheads) and phospho-γ-H2A.X$^+$ (green) microglia (white) 6 weeks post HSCT. Scale bar, 10 µm. **i**, Host microglial Ki67$^+$/phospho-γ-H2A.X$^+$ density (mean ± s.d.; two-sided Kolmogorov–Smirnov test; **$P = 0.0017$; $n = 14$ and 16 brain sections for control and HSCT, respectively). **j**, MCM-2$^+$ (green) and Ki67$^+$ (red) microglia (white, open arrowheads) 6 weeks post HSCT. Scale bar, 10 µm. **k**, Percentage of Ki67$^+$/MCM-2$^+$ cells (mean ± s.e.m.; two-sided, unpaired *t*-test; $P = 0.203$; $n = 5$ animals for control and 6 w HSCT). **l**, PCNA$^+$ (green) and Ki67$^+$ (red) microglia (white, open arrowheads). Scale bar, 10 µm. **m**, Percentage of Ki67$^+$/PCNA$^+$ cells (mean ± s.e.m.; unpaired, two-sided *t*-test; **$P = 0.0011$; $n = 5$ and 6 animals for control and 6 w HSCT, respectively). **n**, pHH3$^+$ (green) and Ki67$^+$ (red) SVZ cells (open arrowheads). Scale bar, 10 µm. **o**, Percentage of Ki67$^+$/pHH3$^+$ cells (mean ± s.e.m.; two-sided Kolmogorov–Smirnov test; **$P = 0.0022$; $n = 5$ and 6 animals for control and 6 w HSCT, respectively).

microglia were observed dividing (Fig. 5f) during 10-min-interval imaging in post-HSCT mice, demonstrating a complete senescent state where they lost their regenerative capacity.

Utilizing the tdTom cell signal in the bloodstream, blood vessels were reconstructed (Supplementary Video 4) and we searched in 3D for transmigration events of tdTom donor cells into the brain parenchyma, but were unable to find any crossings despite tracking in six mice consisting of 27 imaging sessions (10-min intervals) for a total of 28 h in a combined volume of ~0.5 mm^3. The migration front of donor cells was also detected in one mouse at 6 days post PLX, moving at ~40 µm h^{-1} (Extended Data Fig. 7c and Supplementary Video 5) and, by the next day, donor cells filled the imaging field (Extended Data Fig. 7d). At 21 weeks post HSCT, with PLX depletion and recovery, ~91% of the brain was replaced with donor cells, even within white matter, showing the accelerated depletion of host microglia with PLX to be effective in enhancing donor macrophage engraftment[15] (Extended Data Fig. 8a). Interestingly, the striatum was consistently less engrafted, similar to normal post-HSCT engraftment density (Extended Data Fig. 1b), suggesting that the natural microglia turnover rate, which is variable among brain regions[28], probably correlates with engraftment efficiency.

The surveillant process dynamics of donor macrophages in post-PLX HSCT brains were not significantly different from the dynamics of control microglia (Extended Data Fig. 8b), yet post-PLX, HSCT-treated sparse microglia showed decreased dynamics (Extended Data Fig. 8b). There was also no significant difference in neuron density in layers 2/3 between HSCT brains 15 weeks post PLX3397 recovery and control (Extended Data Fig. 8c). Combined, these results indicate that dense donor macrophages may fulfill the role of host microglia and do not have a major

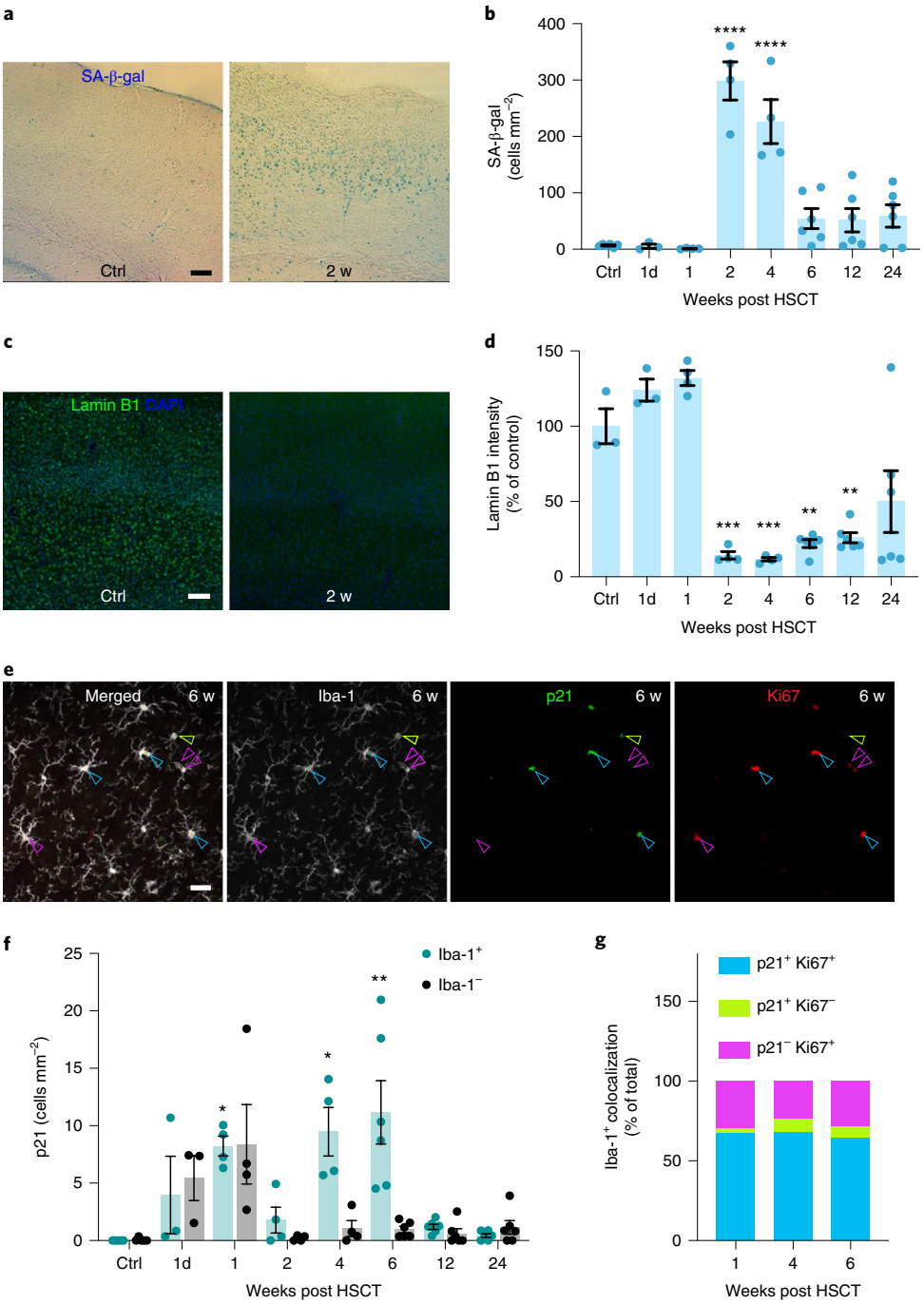


Fig. 4 | Busulfan chemotherapy causes broad senescence in the brain. a, SA-β-gal staining (blue) in control (ctrl) and 2 weeks post HSCT (2 w) dorsal cortex. Scale bar, 100 µm. **b**, Plot of SA-β-gal$^+$ cell density at ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks post HSCT in dorsal cortex (mean ± s.e.m.; one-way, two-sided ANOVA with Tukey's multiple comparisons test; $^{****}P < 0.0001$ versus ctrl; $n = 6$, 3, 4, 4, 4, 6, 6 and 6 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively). **c**, Lamin B1 (green) and DAPI (blue) in ctrl and 2 weeks post HSCT (2 w) in dorsal cortex. Scale bar, 100 µm. **d**, Plot of lamin B1 intensity (percentage of ctrl) at ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks post HSCT in dorsal cortex (mean ± s.e.m.; one-way, two-sided ANOVA with Tukey's multiple comparisons test; $^{***}P < 0.001$, $^{**}P < 0.01$; $P = 0.0008$, $P = 0.0006$, $P = 0.0010$ and $P = 0.0020$ for 2, 4, 6 and 12 weeks, respectively, versus ctrl; $n = 3$, 3, 4, 4, 4, 6, 6 and 6 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively). **e**, Iba-1 (white), p21 (green) and Ki67 (red) labeling showing Iba-1$^+$/p21$^+$/Ki67$^+$ (blue arrowheads), Iba-1$^+$/p21$^-$/Ki67$^+$ (magenta arrowheads) and Iba-1$^+$/p21$^+$/Ki67$^-$ (yellow arrowheads) cells. Scale bar, 30 µm. **f**, p21$^+$ cells that were Iba-1$^+$ or Iba-1$^-$ at ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks post HSCT in dorsal cortex (mean ± s.e.m.; two-sided Kruskal–Wallis with Dunn's multiple comparisons test; $^{**}P < 0.01$, $^{*}P < 0.05$; $P = 0.0306$, $P = 0.0268$ and $P = 0.0045$ for 1, 4 and 6 weeks, respectively, versus ctrl; $n = 5$, 3, 4, 4, 4, 6, 6 and 6 mice for ctrl, 1d and 1, 2, 4, 6, 12 and 24 weeks, respectively in both Iba-1$^+$ and Iba-1$^-$ groups). **g**, Host microglia colabeling combination percentages with p21 and Ki67 at 1, 4 and 6 weeks post HSCT in dorsal cortex (mean percentage of total; one-way, two-sided ANOVA with Tukey's multiple comparisons test; all nonsignificant in comparison of respective labeling combinations at 1 versus 4, 4 versus 6 and 1 versus 6 weeks post HSCT).

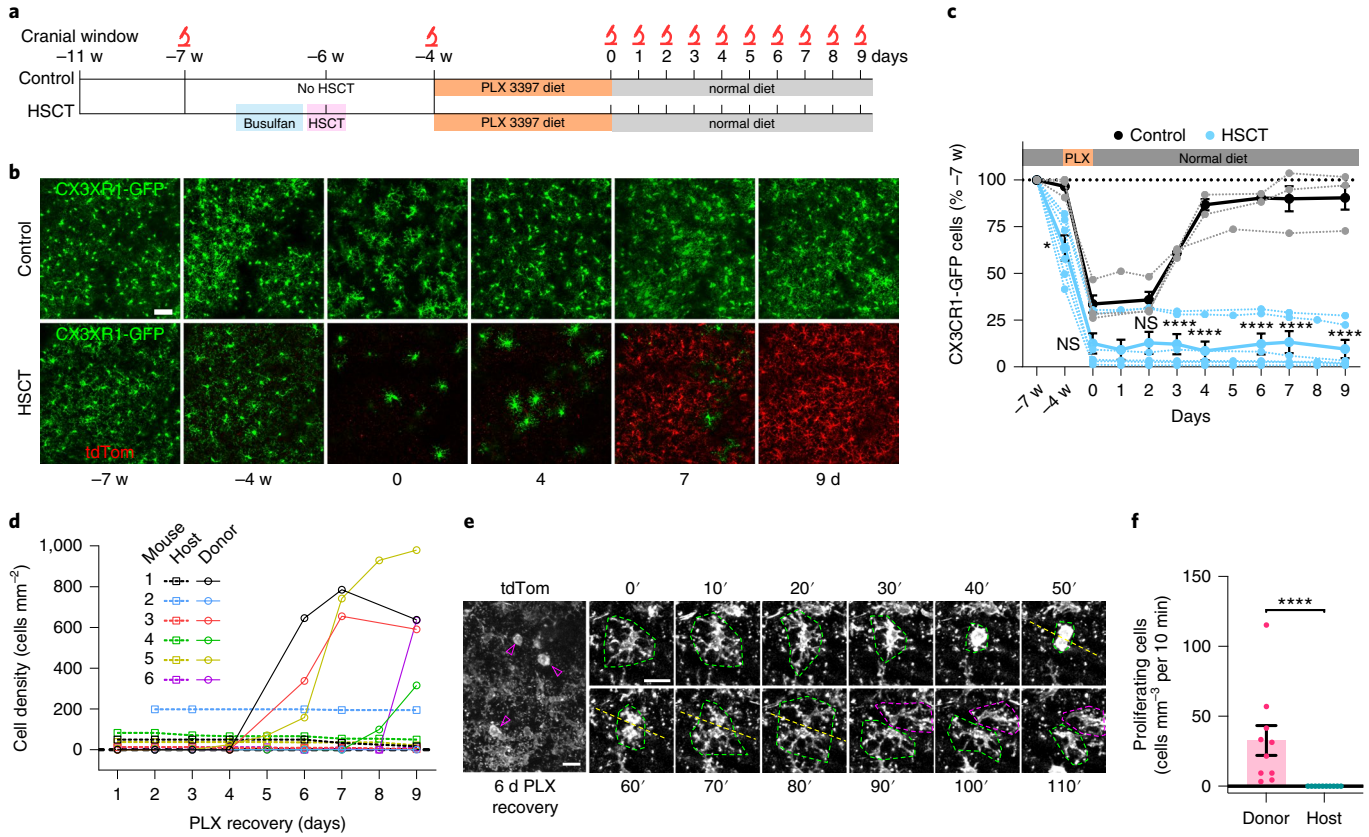


Fig. 5 | HSCT host microglial senescence and incapability to recover from PLX3397-induced depletion cause donor macrophage brain parenchymal engraftment. a, Timeline of in vivo two-photon imaging experiment with control and HSCT groups. The control group underwent neither busulfan treatment nor HSCT, and both groups were placed on a PLX3397 diet for 4 weeks followed by normal diet. Microscope icon indicates in vivo imaging timepoints. **b**, Two-photon in vivo projected images at –7 weeks (–7 w) time point before HSCT, –4 weeks (–4 w) time point before PLX3397 administration and at 0, 4, 7 and 9 days after PLX3397 withdrawal. Control group with CX3CR1-GFP (green) signal showing host microglia. HSCT group showing host microglia (CX3CR1-GFP, green) and subsequent donor macrophage (tdTom, red) engraftment. Scale bar, 50 µm. **c**, Plot of PLX3397 administration and withdrawal effect on control (black, mean ± SEM; gray, individual mice) and HSCT (blue, mean ± s.e.m.; light blue, individual mice) host microglia. One-way, two-sided ANOVA with Tukey multiple comparisons test; $^{*}P = 0.0297$ for –4 weeks and $^{****}P < 0.0001$ for 3, 4, 6, 7 and 9 days post PLX versus respective control time points; $n = 3$ in control and $n = 6$ in HSCT). **d**, Individual plots of each HSCT mouse host microglial (dashed lines, boxes) and donor macrophage (solid lines, open circles) density during PLX3397 recovery. **e**, In vivo two-photon imaged donor macrophage cell division of mouse 3 dorsal cortex 6 days after PLX3397 withdrawal, showing three example cells (arrowheads) undergoing mitosis. Time-lapse images from 0 to 110 min in vivo, imaged at a 10-min intervals, of a donor macrophage (green dashed outline) dividing along its cleavage axis (yellow dashed line) and producing a daughter cell (magenta dashed outline). Scale bar, 20 µm (left) and 5 µm for time series. **f**, Plot of number of in vivo proliferating cells between 5 and 8 days PLX recovery comparison between donor and host (Kolmogorov–Smirnov two-sided test; $^{****}P < 0.0001$; $n = 10$ time-lapse sessions in five mice for both groups).

detrimental effect, supporting a previous study that saw no effect on behavior with macrophage engraftment[11].

Summarizing these results, busulfan chemotherapy HSCT causes complete, irreversible loss of adult neurogenesis, loss of lamin B1 and depletion of half of the host microglial population within 2 weeks post HSCT, which corresponds to increased individual host microglial area, a transient period of Ki67 ramped expression that arrests in parallel with the gradual engraftment of donor macrophages (Fig. 6a). We propose a depletion model that allows the permissive engraftment of peripheral macrophages in response to the partial loss and complete senescence of host microglia (Fig. 6b). Under normal conditions, microglia have a low turnover rate[29,30] with exclusive self-renewal (Fig. 6b, before HSCT). HSCT busulfan chemotherapy depleted the microglial population by 50%, and busulfan-induced DNA damage caused cells to become senescent (Fig. 6b, 2 w post HSCT). Six weeks post HSCT, microglial processes expanded to maintain brain tiling or, due to senescence-induced cell enlargement[31], undergoing cell cycle arrest ($Ki67^{+}/p21^{+}$; Fig. 6b, 6 w post HSCT). A minimal microglial density threshold was reached and the permissive transmigration of peripheral cells ensued, causing their division to re-establish brain tiling (Fig. 6b, 12 w post HSCT), and potentially trophic (interleukin 34)[32] or contact feedback mechanisms prevent host microglia from further initiating cell division.

Discussion

We show that busulfan chemotherapy causes pervasive host microglial senescence and cell cycle arrest, exhausting their regenerative capacity and resulting in brain engraftment of peripheral macrophages, which become resident. Post-HSCT microglia also expand their processes to potentially maintain brain tiling, which is not in accordance with chronic brain inflammation as the engraftment mechanism where an ameboid/condensed microglial morphology would be expected[33]. A previous study demonstrated that parabiosis of a GFP mouse to an inducible CX3CR1-CSF1R-KO mouse, followed by depletion of microglia specifically in the knockout (KO)

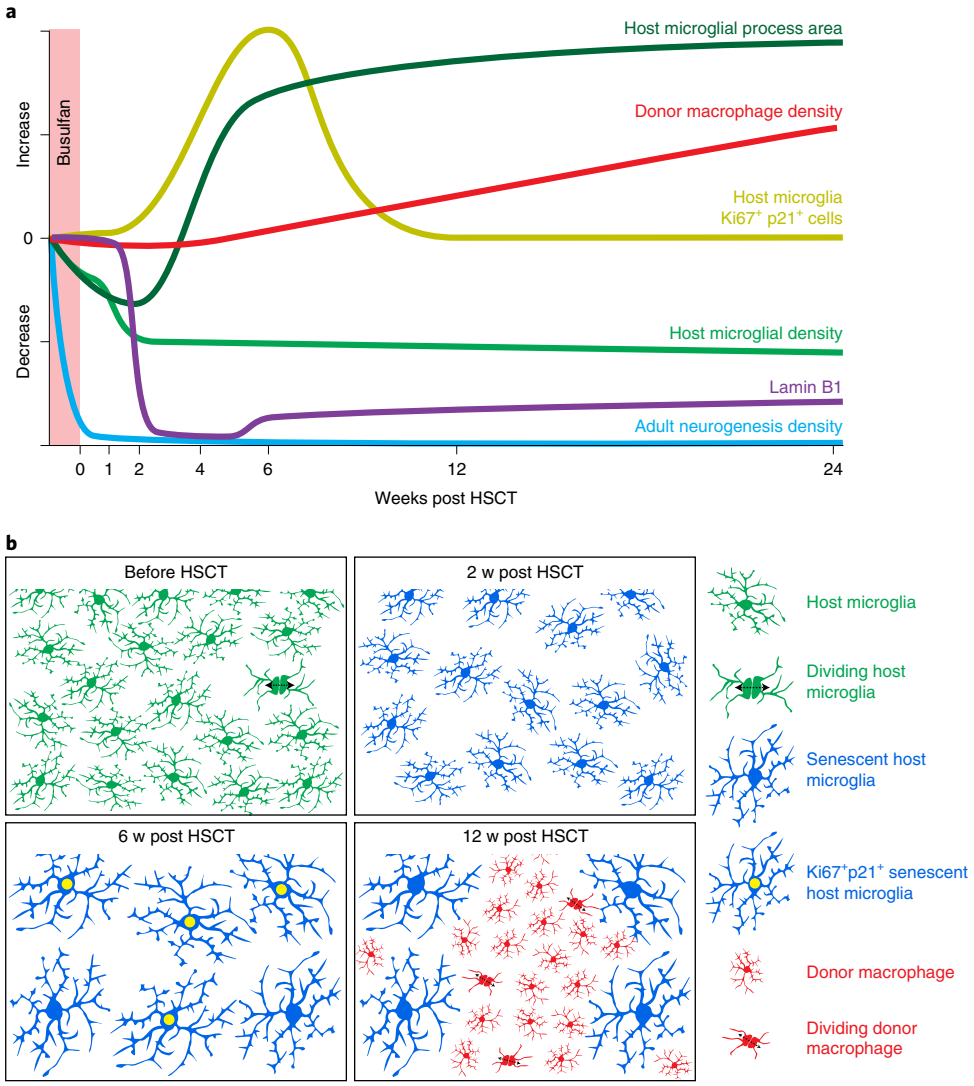


Fig. 6 | Summary of results and busulfan—induced microglial depletion/senescence model of macrophage brain engraftment. a, Summary illustration plot of results showing the effect of HSCT on host microglial density, host microglial area, host microglial Ki67+ cells, donor macrophage density, adult neurogenesis density and lamin B1 up to 24 weeks post HSCT. **b**, Illustration of proposed busulfan-induced host microglial depletion/senescence model driving donor macrophage brain engraftment.

mouse, was sufficient to cause GFP macrophages to engraft the KO brain without the induction of global inflammation[11]. Instead, the brain engraftment mechanism appears to have the similar goal of HSCT myeloablation, where suppression of host hematopoietic stem cells provides a niche for donor cells to engraft the bone marrow. Here we propose that busulfan chemotherapy stops microglial regeneration, thereby causing cell loss that then reaches a critical density, providing a permissive niche for the engraftment of donor macrophages into the brain.

Brain conditioning by either irradiation or chemotherapy was proposed to cause inflammation and BBB disruption and thereby being essential for donor cell engraftment, which was based on head protection during irradiation and a BBB-impermeable chemotherapy agent treosulfan showing no engrafted donor-derived macrophages in the central nervous system after HSCT[7–9]. We instead propose that both these manipulations merely protected the host microglia from irradiation or chemotherapy damage, and thus the population was not depleted/senescent, which consequently deprived donor cells of a permissive niche to engraft the brain. Additionally, busulfan treatment (80 mg kg^{-1}) had no effect on BBB integrity[34], and genetic microglial depletion with transplantation resulted in no change in BBB integrity during macrophage brain engraftment[12]. Taken together, our findings suggest a non-inflammation- or non-BBB-disruption-induced permissive engraftment process, but further direct investigation would be needed.

Our observation that host microglia become senescent after HSCT is confounding, since many cells expressed the commonly used proliferation marker Ki67. Ki67 protein is expressed at the start of G1, being maintained through all subsequent phases of the cell cycle[35]. Nevertheless, Ki67 expression was observed after DNA damage in G_1/S-phase arrested cells[20]. We showed near-complete loss of DNA incorporation in post HSCT microglia, expression of the DNA damage marker pγH2A.X, coexpression of Ki67 with MCM-2 and partially with PCNA, expression of the senescence marker p21 and a complete lack of coexpression with pHH3, a G2 marker. Taken together, we conclude that microglia initiate cell division but that busulfan induces DNA crosslinking[18,36], causing them to become incapable of DNA synthesis and thereby arresting the cell cycle at S phase and becoming senescent via the p21 pathway.

Busulfan also causes hematopoetic stem cell senescence without undergoing apoptosis[16,36], and a similar mechanism may occur with microglia[37]. Interestingly, we and others observed a 50% loss of microglia within the first 2 weeks[38], and it would be intriguing to compare the lost and surviving populations.[27] A similar microglial senescence mechanism may occur with full-body-irradiation myeloablation, since this induces hematopoietic stem cell senescence[36]. Further study on the effect of irradiation on microglial senescence would be important to pursue.

We also discovered busulfan as a chemotherapy agent that rapidly and irreversibly arrests all adult neurogenesis. One day after the 4-day busulfan regimen, almost all cell proliferation was absent in the SVZ and dentate gyrus, accompanied by loss of doublecortin progenitor/immature neurons in the olfactory bulb and hippocampus, persisting up to 24 weeks post HSCT. Previous studies ablated adult neurogenesis to unmask its function in utilizing irradiation[39] and chemotherapy agents, including b-arabinofuranoside[40] and temozolomide[41], yet residual neurogenesis still occurred or recovered with these methods. Clinically, it would be intriguing to determine whether patients who received busulfan chemotherapy lack adult neurogenesis and, moreover, whether this could explain the cognitive deficits associated with HSCT[42].

We also observed a transient loss of NG2 oligodendrocyte precursor cells that, unlike adult neurogenesis and microglia, gradually recovered. It would be interesting to pursue what mechanism protects the NG2 population and whether busulfan treatment causes white matter disruption. This would be particularly important for HSCT treatments in the case of leukodystrophies, since donor cells have been detected in clinical post-mortem brain and spinal cord tissue, and a similar mechanism probably exists in humans[6,43].

We demonstrated in vivo the mechanism of peripheral cell brain engraftment undergoing presumably rare transmigration events followed by massive proliferation and parenchymal migration, albeit after a microglial chemical depletion and recovery protocol. Our inability to observe peripheral transmigration events after recovery from microglial depletion suggests that BBB-crossing events are rare. Furthermore, visualization of migration fronts supported this finding, and future studies with clonal analysis could potentially determine transmigration frequency.

Numerous studies have compared the transcription profile between brain parenchymal macrophages and microglia, with about 90% of transcripts having similar expression[6], yet significant differences were found in the remaining transcripts[6,11,14,15]. When microglia isolated from a donor brain were injected into the hippocampus of post PLX brains, their transcriptome was highly similar to control microglia, suggesting that PLX depletion and recovery had not drastically changed the niche[15]. In contrast, post-HSCT macrophages maintained a more macrophage-like, rather than microglia-like, transcriptome, showing that even near-complete macrophage replacement in the brain environment is not sufficient to change their transcriptome[15]. Although we observed no obvious detrimental effect on cortical neurons with macrophage replacement of microglia, it would be important to determine whether parenchymal macrophages can fulfill the role of microglia, with considerable relevance for future cell therapy strategies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-022-01691-9.

Received: 15 April 2021; Accepted: 10 January 2022;
Published online: 21 February 2022

References

1. Kierdorf, K., Katzmarski, N., Haas, C. A. & Prinz, M. Bone marrow cell recruitment to the brain in the absence of irradiation or parabiosis bias. *PLoS ONE* **8**, e58544 (2013).
2. Ciurea, S. O. & Andersson, B. S. Busulfan in hematopoietic stem cell transplantation. *Biol. Blood Marrow Transplant.* **15**, 523–536 (2009).
3. Aubourg, P. et al. Reversal of early neurologic and neuroradiologic manifestations of X-linked adrenoleukodystrophy by bone marrow transplantation. *New Engl. J. Med.* **322**, 1860–1866 (1990).
4. Cartier, N. et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **326**, 818–823 (2009).
5. Orchard, P. J. & Tolar, J. Transplant outcomes in leukodystrophies. *Semin. Hematol.* **47**, 70–78 (2010).
6. Shemer, A. et al. Engrafted parenchymal brain macrophages differ from microglia in transcriptome, chromatin landscape and response to challenge. *Nat. Commun.* **9**, 5206 (2018).
7. Mildner, A. et al. Microglia in the adult brain arise from Ly-6ChiCCR2^{+} monocytes only under defined host conditions. *Nat. Neurosci.* **10**, 1544–1553 (2007).
8. Peake, K. et al. Bone marrow-derived cell accumulation in the spinal cord is independent of peripheral mobilization in a mouse model of amyotrophic lateral sclerosis. *Front. Neurol.* **8**, 75 (2017).
9. Zhan, L. et al. Proximal recolonization by self-renewing microglia re-establishes microglial homeostasis in the adult mouse brain. *PLoS Biol.* **17**, e3000134 (2019).
10. Böttcher, C., Fernández-Klett, F., Gladow, N., Rolfes, S. & Priller, J. Targeting myeloid cells to the brain using non-myeloablative conditioning. *PLoS ONE* **8**, e80260 (2013).
11. Cronk, J. C. et al. Peripherally derived macrophages can engraft the brain independent of irradiation and maintain an identity distinct from microglia. *J. Exp. Med.* **215**, 1627–1647 (2018).
12. Lund, H. et al. Competitive repopulation of an empty microglial niche yields functionally distinct subsets of microglia-like cells. *Nat. Commun.* **9**, 4845 (2018).
13. Peake, K. et al. Busulfan as a myelosuppressive agent for generating stable high-level bone marrow chimerism in mice. *J. Vis. Exp.* **1**, e52553 (2015).
14. Butovsky, O. et al. Identification of a unique TGF-β dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
15. Xu, Z. et al. Efficient strategies for microglia replacement in the central nervous system. *Cell Rep.* **32**, 108041 (2020).
16. Meng, A., Wang, Y., Brown, S. A., Zant, G. V. & Zhou, D. Ionizing radiation and busulfan inhibit murine bone marrow cell hematopoietic function via apoptosis-dependent and -independent mechanisms. *Exp. Hematol.* **31**, 1348–1356 (2003).
17. Probin, V., Wang, Y., Bai, A. & Zhou, D. Busulfan selectively induces cellular senescence but not apoptosis in WI38 fibroblasts via a p53-independent but extracellular signal-regulated kinase-p38 mitogen-activated protein kinase-dependent mechanism. *J. Pharmacol. Exp. Ther.* **319**, 551–560 (2006).
18. Iwamoto, T. et al. DNA intrastrand cross-link at the 5'-GA-3' sequence formed by busulfan and its role in the cytotoxic effect. *Cancer Sci.* **95**, 454–458 (2004).
19. Abdallah, N. M.-B. B., Slomianka, L., Vyssotski, A. L. & Lipp, H.-P. Early age-related changes in adult hippocampal neurogenesis in C57 mice. *Neurobiol. Aging* **31**, 151–161 (2010).
20. Oijen, M. G. C. T., van, Medema, R. H., Slootweg, P. J. & Rijksen, G. Positivity of the proliferation marker Ki-67 in noncycling cells. *Am. J. Clin. Pathol.* **110**, 24–31 (1998).
21. Lombardi, D., Becherucci, F. & Romagnani, P. How much can the tubule regenerate and who does it? An open question. *Nephrol. Dial. Transplant.* **31**, 1243–1250 (2016).
22. Sasaki, K., Matsumura, K., Tsuji, T., Shinozaki, F. & Takahashi, M. Relationship between labeling indices of KI-67 and brdurd in human malignant tumors. *Cancer* **62**, 989–993 (1988).
23. Lee, B. Y. et al. Senescence-associated β-galactosidase is lysosomal β-galactosidase. *Aging Cell* **5**, 187–195 (2006).
24. Freund, A., Laberge, R.-M., Demaria, M. & Campisi, J. Lamin B1 loss is a senescence-associated biomarker. *Mol. Biol. Cell* **23**, 2066–2075 (2012).
25. Barr, A. R. et al. DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nat. Commun.* **8**, 14728 (2017).
26. Huang, Y. et al. Repopulated microglia are solely derived from the proliferation of residual microglia after acute depletion. *Nat. Neurosci.* **21**, 530–540 (2018).
27. Elmore, M. R. P. et al. Colony-stimulating factor 1 receptor signaling is necessary for microglia viability, unmasking a microglia progenitor cell in the adult brain. *Neuron* **82**, 380–397 (2014).
28. Tay, T. L. et al. A new fate mapping system reveals context-dependent random or clonal expansion of microglia. *Nat. Neurosci.* **20**, 793–803 (2017).

29. Réu, P. et al. The lifespan and turnover of microglia in the human brain. *Cell Rep.* **20**, 779–784 (2017).
30. Askew, K. et al. Coupled proliferation and apoptosis maintain the rapid turnover of microglia in the adult brain. *Cell Rep.* **18**, 391–405 (2017).
31. Campisi, J. & Fagagna, Fd'Addadi Cellular senescence: when bad things happen to good cells. *Nat. Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
32. Oosterhof, N. et al. Colony-stimulating factor 1 receptor (CSF1R) regulates microglia density and distribution, but not microglia differentiation in vivo. *Cell Rep.* **24**, 1203–1217 (2018).
33. Jonas, R. A. et al. The spider effect: morphological and orienting classification of microglia in response to stimuli in vivo. *PLoS ONE* **7**, e30763–12 (2012).
34. Lampron, A., Lessard, M. & Rivest, S. Effects of myeloablation, peripheral chimerism, and whole-body irradiation on the entry of bone marrow-derived cells into the brain. *Cell Transplant.* **21**, 1149–1159 (2012).
35. Scholzen, T. & Gerdes, J. The Ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* **182**, 311–322 (2000).
36. Meng, A., Wang, Y., Zant, G. V. & Zhou, D. Ionizing radiation and busulfan induce premature senescence in murine bone marrow hematopoietic cells. *Cancer Res.* **63**, 5414–5419 (2003).
37. Jänicke, R. U., Sohn, D., Essmann, F. & Schulze-Osthoff, K. The multiple battles fought by anti-apoptotic p21. *Cell Cycle* **6**, 407–413 (2007).
38. Capotondo, A. et al. Brain conditioning is instrumental for successful microglia reconstitution following hematopoietic stem cell transplantation. *Proc. Natl Acad. Sci. USA* **109**, 15018–15023 (2012).
39. Valley, M. T., Mullen, T. R., Schultz, L. C., Sagdullaev, B. T. & Firestein, S. Ablation of mouse adult neurogenesis alters olfactory bulb structure and olfactory fear conditioning. *Front. Neurosci.* **3**, 51 (2009).
40. Ghanbari, A. et al. Depletion of neural stem cells from the subventricular zone of adult mouse brain using cytosine b-arabinofuranoside. *Brain Behav.* **5**, e00404 (2015).
41. Egeland, M. et al. Depletion of adult neurogenesis using the chemotherapy drug temozolomide in mice induces behavioural and biological changes relevant to depression. *Transl. Psychiatry* **7**, e1101 (2017).
42. Buchbinder, D. et al. Neurocognitive dysfunction in hematopoietic cell transplant recipients: expert review from the late effects and Quality of Life Working Committee of the CIBMTR and complications and Quality of Life Working Party of the EBMT. *Bone Marrow Transplant.* **53**, 535–555 (2018).
43. Wolf, N. I. et al. Metachromatic leukodystrophy and transplantation: remyelination, no cross-correction. *Ann. Clin. Transl. Neurol.* **7**, 169–180 (2020).

Methods

HSCT. All experimental procedures followed approved guidelines regulated by the institutional and national committees (Comité d'éthique en expérimentation animale no. 89). Donor mice (male, average age 25.8 ± 2.0 weeks) endogenously expressing tdTomato[44] (B6.129(Cg)-Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$; Jackson Laboratory stock no. 007676) were sacrificed and their lower limbs dissected to obtain femora and tibiae. Cells were extracted from bone marrow by infusion of medium (IMDM; Fisher, no. 11504556, 1% Pen/Strep), aggregates were manually dissociated and the suspension was strained (40 µm). Cells were counted and diluted for lineage-negative purification using the Mouse Lineage Cell Depletion Kit (Miltenyi Biotec, no. 130-090-858). After purification following the manufacturer's protocol, Lin$^-$ cells were maintained at 37 °C (5% CO_2) for 24 h (IMDM, 10% fetal bovine serum; mSCF 50 ng ml^{-1}; hIL6, 50 ng ml^{-1};and mIL3, 20 ng ml^{-1}; Peprotech). Myeloablation of recipient mice was performed by the administration of busulfan (25 mg kg^{-1} d^{-1} i.p. for 4 days; total dose, 100 mg kg^{-1}) before HSCT. Recipient male C57BL/6 J adult mice (average age, 14.7 ± 2.7 weeks) were used for histological studies, and CX3CR1-GFP male hemizygous transgenic mice[45] on a C57BL/6 J background (B6.129P2(Cg)-Cx3cr1^{GFP}; Jackson Laboratory stock no. 005582; average age 12.1 ± 0.7 weeks) were used for in vivo two-photon imaging studies. Each mouse was transplanted with ~10^6 cells by retro-orbital injection while under isoflurane anesthesia, and antibiotic (sulfamethoxypyridazine/trimethoprime) was added to the drinking water for the following 7 days. For CSF1R drug administration, PLX3397 (Plexicon, Chemgood) in the mouse diet (290 mg kg^{-1} PLX3397 in U8978 diet; SAFE Diets) was added to cages while control groups received food without drug (U8978 diet; SAFE Diets). All mice were provided normal diet during the PLX3397 withdrawal period.

In vivo two-photon Imaging. For in vivo two-photon imaging studies, cranial window surgery was performed no later than 4 weeks before transplantation, as previously described[46]. The animals were anesthetized (100 mg kg^{-1} ketamine/10 mg kg^{-1} xylazine) and a 5-mm cranial window created over the right dorsal cortex between bregma and lambda sutures. Dental cement was used to seal both the coverglass and a stainless steel head bar to the skull[47]. For each imaging session, mice were lightly anesthetized with isoflurane (0.8%) and the metal head bar was attached to a custom stage. Anesthesia continued throughout imaging, and body temperature was maintained with a rectal thermometer feedback heating pad. A fiducial point of reference on the stage was used to zero *x*–*y*–*z* positions to find the same brain volume over multiple sessions. Imaging was performed on an Investigator microscope (Prairie View/Bruker) with a ×16/0.8 numerical aperture objective (Olympus) and a Ti:Sapphire femtosecond pulsed two-photon laser (DeepSee, Spectra Physics) tuned to 1,000 nm. Multiple *Z*-plane stack images were acquired at a 2–3-µm *z*-step at a resolution of either 0.82 × 0.82^2 or 0.41 × 0.41 µm^2 per pixel *X*–*Y*.

Immunohistochemistry. A cohort of animals at 6 weeks post HSCT was injected with EdU (Sigma-Aldrich, no. 900584; 50 mg kg^{-1} i.p.) 2 days before sacrifice. Animals were anesthetized with a 2× dose of ketamine/xylazine and transcardially perfused with heparinized saline followed by 4% paraformaldehyde. Tissues were dissected and post-fixed in paraformaldehyde for 2–4 h, followed by 24 h in 30% sucrose in PBS. Tissues were microtome sectioned at 40 µm and stored at −20 °C in antifreeze solution. For immunolabeling, brain sections were washed in PBS, underwent antigen retrieval (30 min at 80 °C; citrate buffer pH 6.0), washed in PBS and incubated in various combinations of the following primary antibodies: Iba-1 (1:500, guinea pig; Synaptic Systems, no. 234004), TMEM119 (1:250, rabbit; Abcam, no. ab209064), Ki67 (1:250, rat; Invitrogen, no. 14-5698-80), doublecortin (1:500, guinea pig; Millipore, no. AB2253), NG2 (1:400, rabbit; Chemicon, no. AB5320), phospho-gamma-H2A.X (1:1,000, rabbit,; Abcam, no. ab2893), MCM-2 (1:250, mouse; BD Biosciences, no. 610701), PCNA (1:250, rabbit; Abcam, no. EPR3821), pHH3 (1:500, rabbit; Invitrogen, no. PA5-104936), lamin B1 (1:500, rabbit; Abcam, no. AB16048), NeuN (1:500, mouse; Millipore, no. MAB377) and p21 (1:250, rabbit; Abcam, no. AB188224). Primary antibody was diluted in PBS^{++} (PBS, 5% normal donkey serum, 0.25% triton) and incubated overnight at 4 °C. Sections were then washed in PBS and incubated for 2 h at room temperature with combinations of the following secondary antibodies in PBS^{++}: donkey anti-guinea pig Alexa Fluor 648 (1:500; Jackson Immno Research, no. 706-605-148), goat anti-rabbit Alexa Fluor 568 (1:500; Invitrogen, no. A11036), goat anti-rabbit Alexa Fluor 488 (1:500; Invitrogen, no. A11034), goat anti-rabbit Alexa Fluor 647 (1:500; Invitrogen, no. A21244), goat anti-rat Alexa Fluor 568 (1:500; Invitrogen, no. A11077) and goat anti-mouse Alexa Fluor 586 (1:500; Invitrogen, no. A11031). Sections were then washed in PBS. EdU was labeled with an Alexa Fluor 488 Click-iT cell proliferation kit following the manufacturer's protocol (ThermoFisher, no. C10337). Fluorescent sections were mounted, coverslipped and confocal imaged (Zeiss LSM 700, Zen Black software) at a resolution of 0.63 × 0.63 µm^2 per pixel for image analysis. Brain sections were incubated at 37 °C overnight for SA-β-gal staining (Cell biolabs, no. CBA-230) using the manufacturer's protocol. Sections were mounted, dehydrated and coverslipped, with brightfield images acquired using a Zeiss Observer Z1 microscope with a Hamamatsu ORCA-ER camera at a resolution of 0.65 × 0.65 µm^2 per pixel.

Image analysis. For in vivo two-photon data, 3D registration was performed with 3D slicer software v.4.10.2 (www.slicer.org) and/or Imaris software (Bitplane, v.9.1.2). Images were normalized across time points and *z*-planes using Matlab 2018a (Mathworks) code[48]. Donor macrophage 3D surface renderings were manually formed using the magic wand tool in the surface toolkit. For the PLX3397 depletion and recovery time series, supervised spot analysis in Imaris was used for tracking of individual cells across days. Cells that were within 40 µm of the *x*, *y* or *z* volume edges were excluded from analysis, to eliminate migration- or registration cropping-associated loss/gain artifacts. Blood vessel reconstruction took advantage of the tdTom signal in circulating blood cells. A simple horizontal spatial filter was applied to images due to the microscope *x*-scan speed being greater than on the *y* axis, and therefore cells in the bloodstream were *x*-distorted from their rapid trajectory. Multiple timepoints of extracted cells from time-lapse imaging were registered and summed to reconstruct the intraluminal vasculature structure. To search for peripheral cell transmigration events, in Imaris software using 3D glasses (Nvidia 3dVision) cells were followed within the bloodstream in a 10-min-interval image series, searching for their transmigration into the parenchyma (each imaging session was between four and ten volumes). Quantification of host microglia and donor macrophage confocal images was performed with Fiji ImageJ (imagej.net/Fiji, NIH, v.1.53) software. Confocal images were indexed with a number upon acquisition, with the observer performing subsequent quantification blind to the image reference. Images were cropped to exclude large blood vessels and cortical edges. The image was split to show overlays of all fluorescent channels separately, and the 'synchronize windows' tool was used to show the pointer location in all images. Using the 'multi-point' tool, cell types were manually counted based on the combination of antibody markers. For each brain section, convex hull circumference was manually traced and measured for five random cells using the 'polygon selection' tool. Donor macrophage and host microglial process motility was recorded in vivo by acquisition of image stacks every 10 min. In Imaris, a 70-µm projection was made for each field of view (FOV) at each time point, the series was normalized and registered ('image stabilizer';https://www.cs.cmu.edu/~kangli/code/Image_Stabilizer.html) and five cells from each FOV with complete structure underwent convex hull tracing of their process area. In Matlab, a contiguous spatial filter of ten pixels was applied and iterative subtraction of images within the series ($(n+1)-n$) was performed for detection of motility. The subtraction product images were binarized, summed and divided by the area (pixels) of the complete cell projection to calculate percentage change per 10 min.

Statistical analysis. Plots and statistics were performed with Prism software v.9 (www.graphpad.com). A Shapiro–Wilk normality test was first performed and, for normal data, one-way two-sided analysis of variance (ANOVA) with Tukey's multiple comparisons test was performed; otherwise, for non-normal data, a two-sided Kruskal–Wallis or Kolmogorov–Smirnov test was performed. For microglial process area *x*–*y* plots, a simple linear regression was performed with the extra sum-of-squared *F*-test to determine whether the fit was significant from the null hypothesis of slope = 0.

Statistics and reproducibility. Representative images were examples from quantified data in the following pairs ('Image'–'data plot'): 1c–1d, 1e–1f, 1g–1h, 2b–2a and –2c, 3a–3b, 3c–3d, 3f–3g, 3h–3i, 3j–3k, 3l–3m, 3n–3o, 4a–4b, 4c–4d, 4e–4f and –4 g, 5b–5c and –5d, 5e–5f; and Extended Data Figs. 1c–1d, 2a–2b, 2c–2d, 3a–3b, 3c–3e, 3d–3g, 4a–4b and 4c–4b. These image–data plot pairs have the number of animals/replicates per group indicated in the data plot figure legend. For histology confocal images, each animal had at least four replicates by quantification of two brain sections on each hemisphere. Figure 2d is a representative image without a data plot, but Fig. 5a and Extended Data Figs. 7a and 8a show the same complementary tiling pattern between microglia and donor macrophages. Extended Data Fig. 7c is derived from a 130-min-interval imaging session of the same region, and Extended Data Fig. 7d of that region across three separate days in the same animal.

Microglial Ki67 density age-decline neurogenesis modeling. Statistical analysis and modeling were performed in Python, and goodness of fit was assessed with R^2 and root mean square error (RMSE). Since Ki67 is expressed for ~32 h in microglia[30], microglial proliferation rate was fit over time based on the mean of measured data for each time point. Spline interpolation of order 3 was used, meaning that the first-order derivative is null on the first and sixth points (weeks 0 and 12), to ensure that the proliferation rate remained positive ($R^2 = 0.74$ and RMSE = 2.64). Microglial density (cells mm^{-2}) was fit over the first five timepoints, until week 6. It is assumed that microglia divide in two separate populations: N_1, the resident microglia and N_2, newly generated microglia. A distinct rate model was used to fit each population evolution, the rate of death being λ_1 and λ_2 for N_1 and N_2, respectively; $\mu(t)$, the proliferation rate for N_2, was interpolated. N_1: $\frac{dN_1}{dt}(t) = -\lambda_1 N_1(t)$ leading to $N_1(t) = N_1^0 \times e^{-\lambda_1 t}$. N_2: $\frac{dN_2}{dt}(t) = \mu(t) - \lambda_2 N_2(t)$, which was solved iteratively using Euler's method, with a time step (dt) of 1 day: $N_2(t + dt) = N_2(t) \times (1 - \lambda_2)dt + \mu(t) \times \mathrm{dt}$. To fit total microglial density $N(t) = N_1(t) + N_2(t)$, we optimized successively λ_1 and λ_2 using the least-squares method, obtaining $\lambda_1 = 5.1\%$/day, $\lambda_2 = 10.8\%$/day and $N_1^0 = 281.8$ ($R^2 = 0.84$

and RMSE = 27.46). To fit our control neurogenesis Ki67$^+$ (Fig. 3b) and DCX$^+$ (Extended Data Fig. 3e) cell quantification to previously published, age-related decline results[19], data were fitted to a distinct rate model with N being the population and λ the corresponding monthly death rate: $\frac{dN}{dt}(t) = -\lambda N(t)$ leading to $N(t) = N^0 \times e^{-\lambda t}$. For Ki67$^+$cells, $\lambda = 39.7\%$/day and $N^0 = 8{,}250$ ($R^2 = 0.91$ and RMSE = 642); and for DCX$^+$ cells, $\lambda = 33.7\%$/day and $N^0 = 28{,}566$ ($R^2 = 0.89$ and RMSE = 2,435).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data used in the figures for this study are provided in Supplementary information and are covered by the Creative Commons Attribution 4.0 International License (no. CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

References

44. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
45. Jung, S. et al. Analysis of fractalkine receptor CX3CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell Biol.* **20**, 4106–4114 (2000).
46. Holtmaat, A. et al. Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window. *Nat. Protoc.* **4**, 1128–1144 (2009).
47. Sailor, K. A. et al. Persistent structural plasticity optimizes sensory information processing in the olfactory bulb. *Neuron* **91**, 384–396 (2016).
48. Sun, G. J. et al. Seamless reconstruction of intact adult-born neurons by serial end-block imaging reveals complex axonal guidance and development in the adult hippocampus. *J. Neurosci.* **33**, 11400–11411 (2013).

Acknowledgements

We thank P. Aubourg for his inspiration to initiate this study; J. Strauch, I. Bernstein and G. Dufayet-Chaffaud for experimental support; E. Gomez Perdiguero, G. Lepousez, I. Gabanyi, S. Voytek, F. Koukouli, H. Song, G. Eberl, L. Peduto, F. Jagot-Brunner, G. Parsons and A. Giniatullina for advice; and T. Sailor for two-photon stage fabrication. This work was partly funded by bluebird bio, Inc. as part of a collaboration (K.A.S., P.-M.L. and N.C.), Agence Nationale de la Recherche (no. ANR-15-CE37-0004-01, to K.A.S. and P.-M.L.), Agence Nationale de la Recherche (no. ANR-15-NEUC-0004, Circuit-OPL, to K.A.S. and P.-M.L.), the life insurance company 'AG2R-La-Mondiale' (to K.A.S. and P.-M.L.), the Agence Nationale de la Recherche 'Investissements d'avenir' program (no. ANR-19$^-$P3IA-0001, PRAIRIE 31A Institute, to C.G., J.-B.M. and C.L.V.) and a pre- and postdoctoral fellowship from the Laboratory for Excellence (LabEx) 'Revive' (no. ANR-10-LABX-73, to C.G. and K.A.S.). S.L.-M, S.T., B.G.-L. and N.C. funding from the program "Investissements d'avenir" ANR-10-IAIHU-06 and ANR-11-INBS-0011–NeurATRIS: Translational Research Infrastructure for Biotherapies in Neurosciences. Part of this work was carried out in the PHENOPARC core facility, CELIS core facility and ICAN core facility of ICM.

Author contributions

K.A.S. undertook conceptualization, investigation, methodology and writing of the original draft. G.A. carried out conceptualization, investigation, methodology, writing review and editing. S.L.-M., B.G.-L. and S.T. were responsible for conceptualization and investigation. C.G. and C.L.V. carried out methodology. J.-B.M. performed methodology and supervision. M.B. and K.G. were responsible for conceptualization, writing of the review and editing. G.V. undertook conceptualization. P.-M.L. carried out supervision, conceptualization, writing of the review and editing. N.C. carried out supervision, conceptualization, writing of the review and editing.

Competing interests

This project was partially funded by bluebird bio, Inc. for reagents and supplies, with partial salaries provided to G.A., B.G.-L., S.L.-M. and S.T.; M.B. and K.G. are current employees, and G.V. a former employee, of bluebird bio, Inc.; K.A.S., C.G., J.-B.M., C.L.V., P.-M.L. and N.C. declare no competing interests.

Additional information

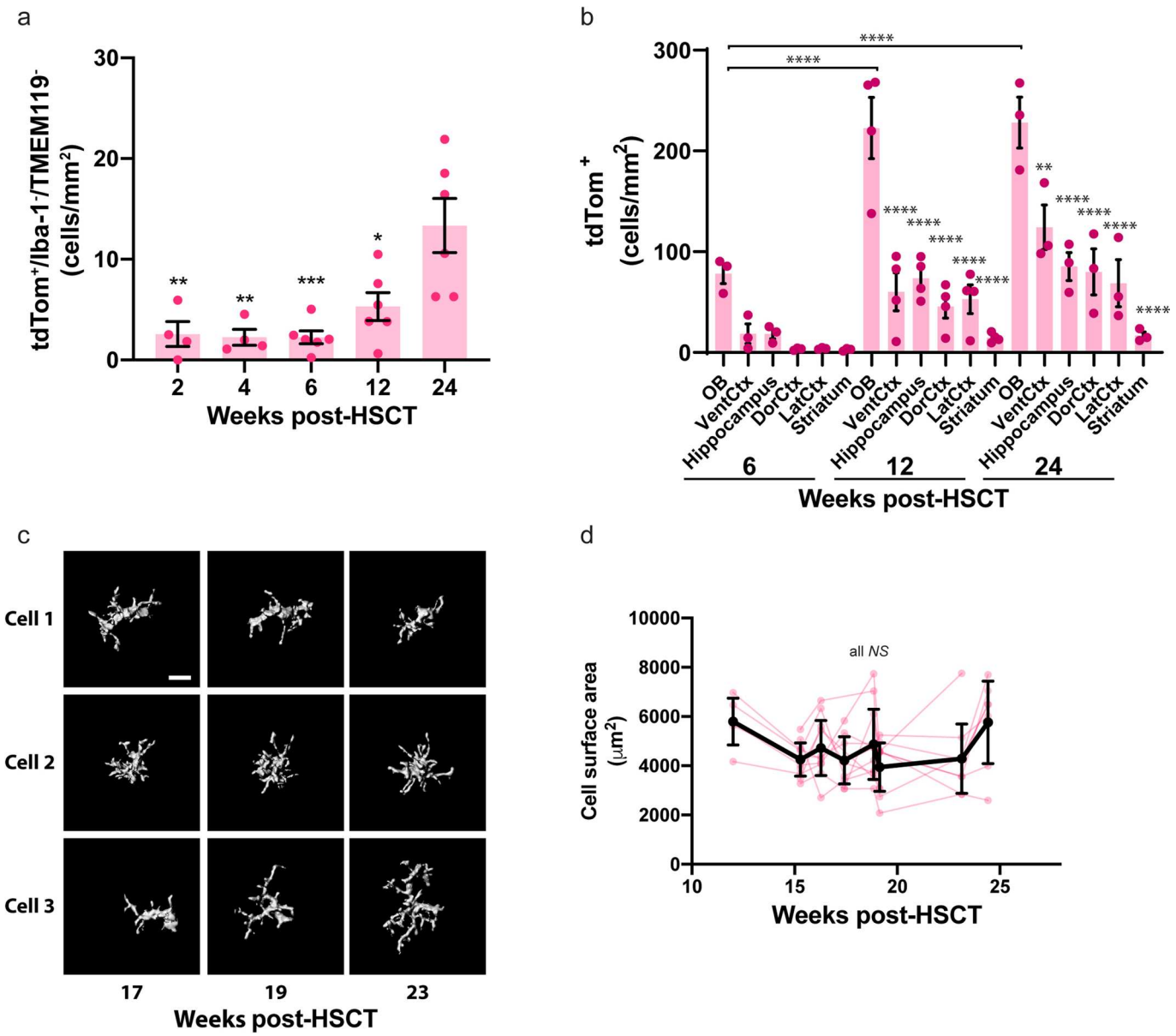
Extended data is available for this paper at https://doi.org/10.1038/s41591-022-01691-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-022-01691-9.

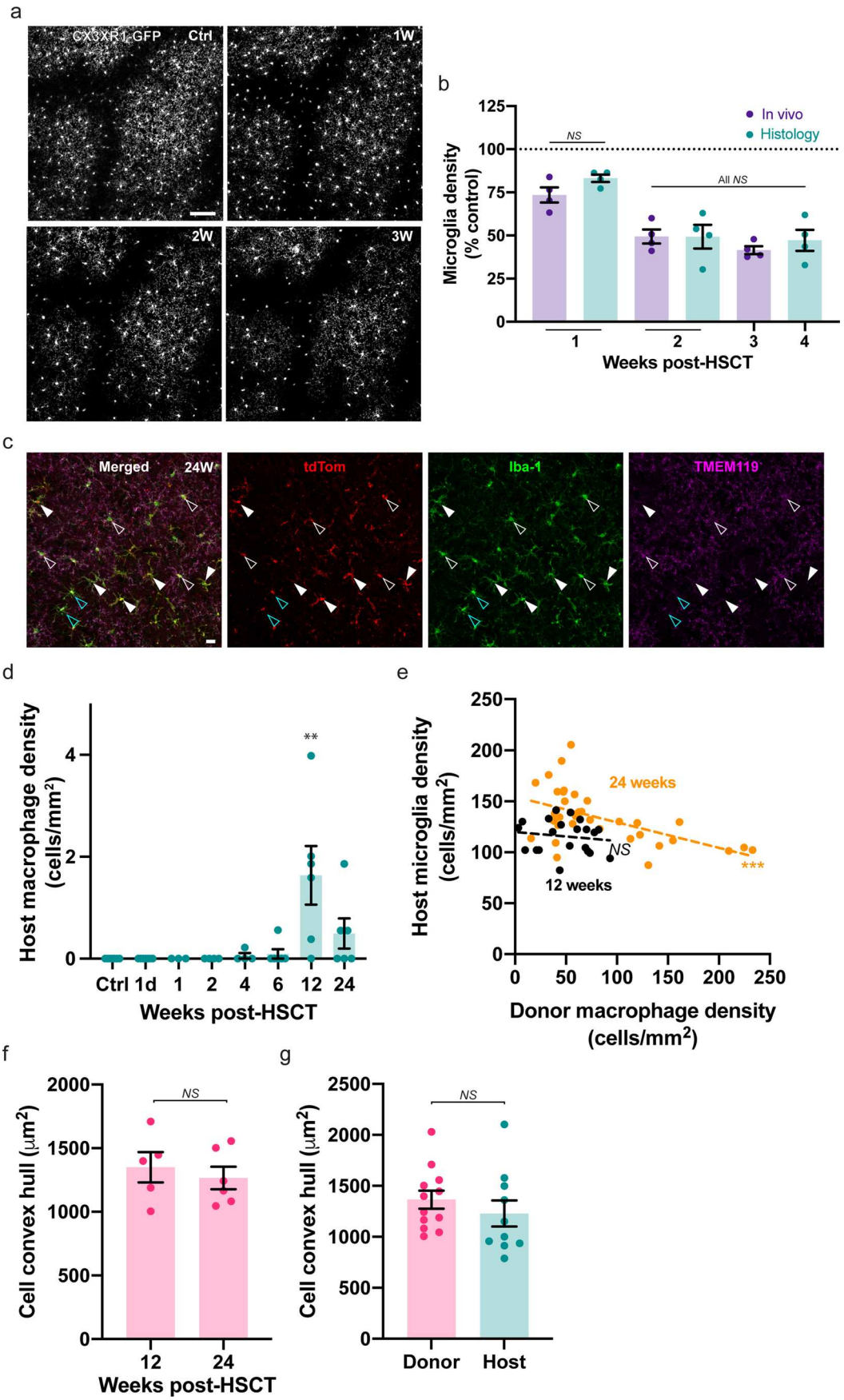
Correspondence and requests for materials should be addressed to Kurt A. Sailor, Pierre-Marie Lledo or Nathalie Cartier.

Peer review information *Nature Medicine* thanks Joanne Kurtzberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Jerome Staal was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team

Reprints and permissions information is available at www.nature.com/reprints.

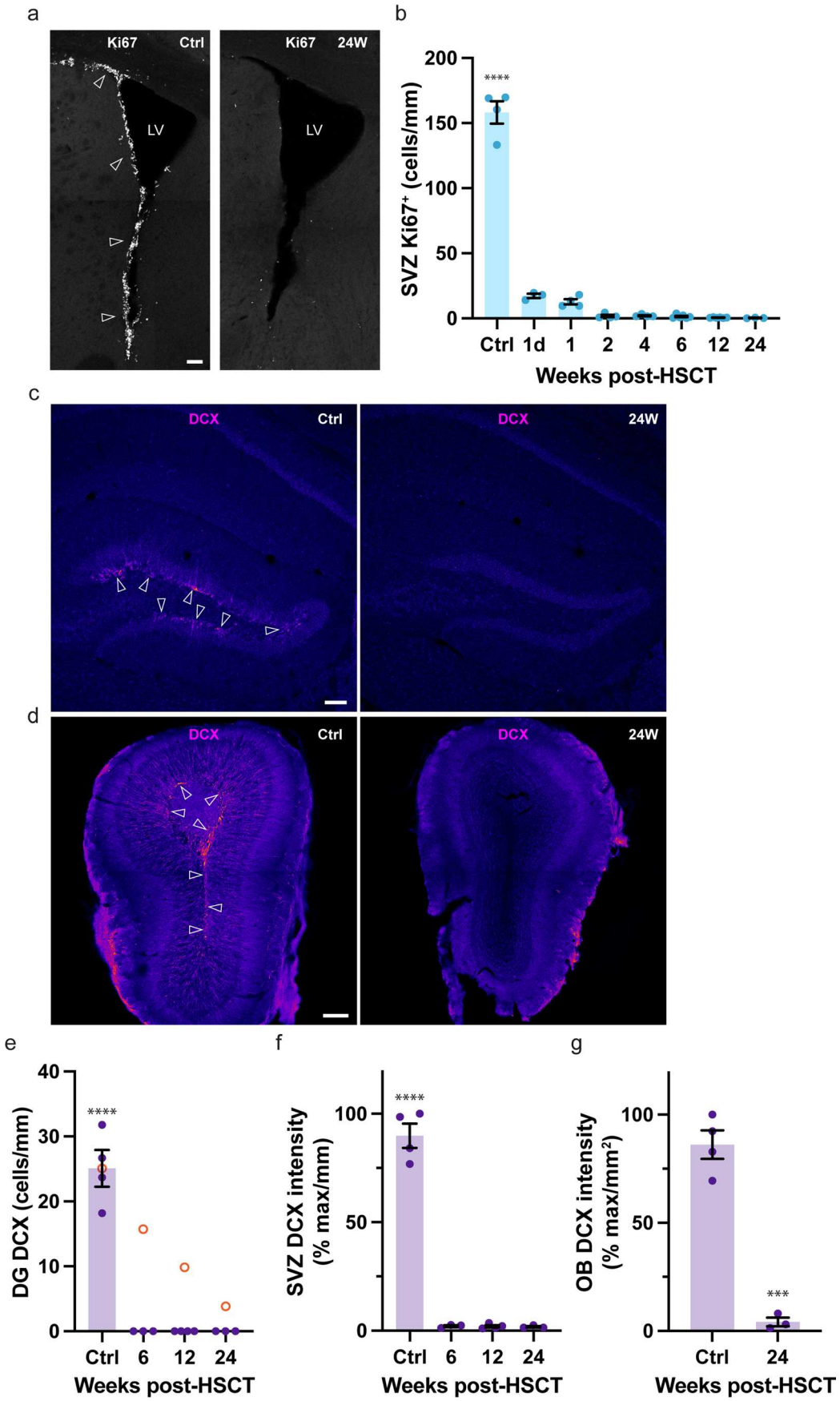


Extended Data Fig. 1 | HSCT donor cell macrophage characteristics within brain. (a) Plot of donor cells that are negative for macrophage or microglia markers (tdTom$^+$/Iba-1$^-$/TMEM119$^-$; mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; $P = 0.0025$, 0.0019, 0.0006 and 0.0128 for 2, 4, 6 12 and 24 weeks, respectively, vs. 24 weeks; n = 4 mice for 2 and 4 weeks, n = 6 mice for 6, 12 and 24 weeks). (b) Plot of density of donor cells (tdTom$^+$) in brain regions of olfactory bulb (OB), hippocampus, dorsal cerebral cortex (DorCtx), lateral cerebral cortex (LatCtx) and striatum at 6, 12 and 24 weeks post-HSCT. The OB has a significantly higher density of cells compared to all other brain regions at 12 and 24 weeks. (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; $P < 0.0001$ for OB 12 and 24 weeks vs. 6 weeks, ****$P < 0.0001$ for non-OB regions vs. OB at respective week with **$P = 0.0095$ 24 weeks VentCtx vs. OB.; n = 3, 4 and 3 mice for 6, 12 and 24 weeks post-HSCT, respectively). (c) Surface renderings of three example donor cells at 17, 19 and 23 weeks post-HSCT. Scale bar = 15 μm. (d) Cell surface area plot of the same donor cells *in vivo* imaged from 12 to 25 weeks post-HSCT. (Light red plot of individual cells, black plot mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; n = 5, 10, 11, 9, 11, 11, 9, 7 cells for 12, 15.3, 16.3, 17.4, 18.9, 19.1, 23.1 and 24.4 weeks post-HSCT, respectively).



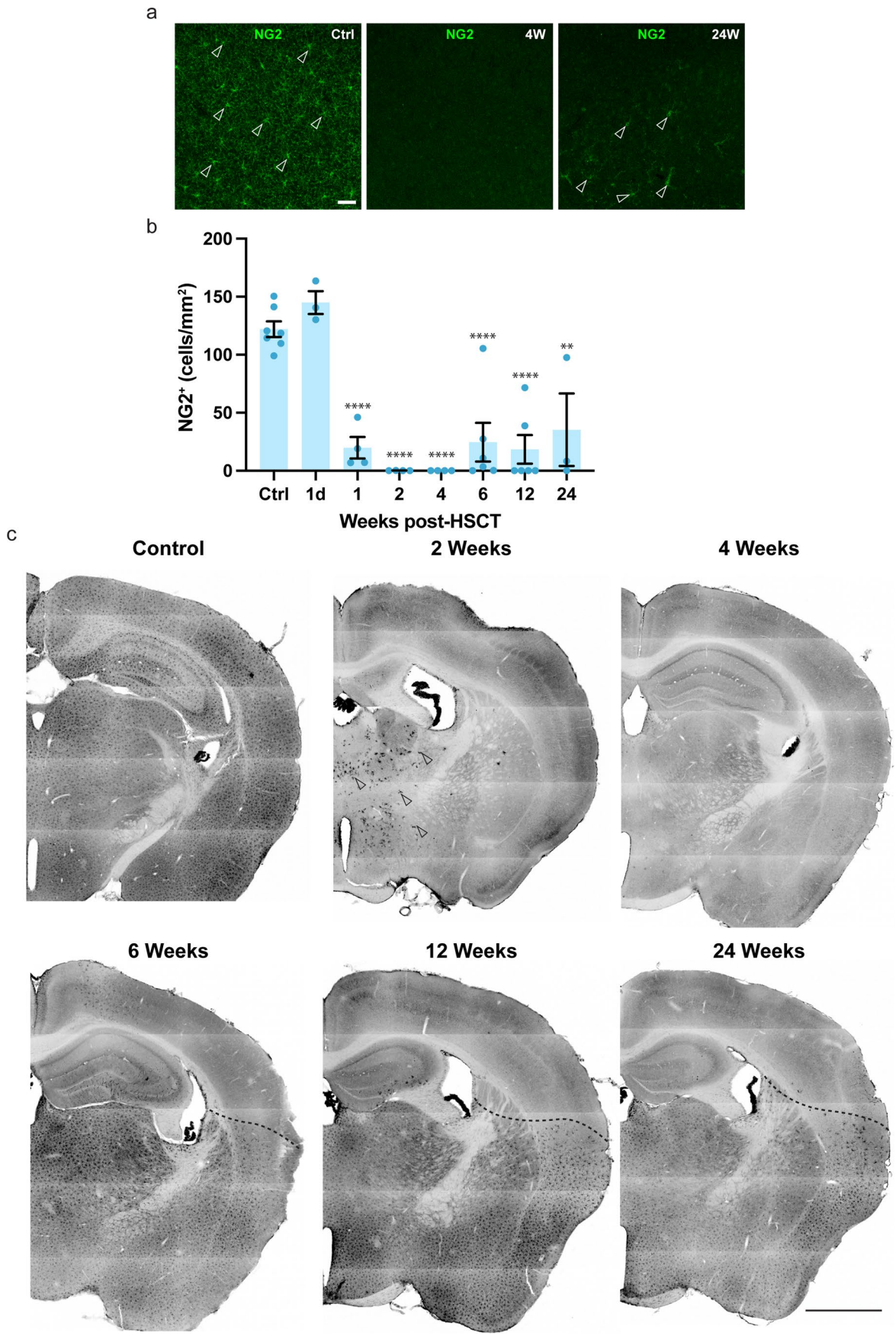
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Host microglia density decline with HSCT, donor and host macrophage properties. (a) The same registered brain region 2-photon *in vivo* imaged. Scale bar = 100 µm. (b) Microglia density comparing between *in vivo* 2-photon imaged CX3CR1-GFP mice and histology ($tdTom^-/Iba\text{-}1^+/TMEM119^+$) analysis. (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; n = 4 animals all groups). (c) Immunohistochemistry at 24 weeks showing donor macrophages ($tdTom^+/Iba\text{-}1^+/TMEM119^-$, white closed arrowhead), host microglia ($tdTom^-/Iba\text{-}1^+/TMEM119^+$, white open arrowhead) and host macrophages ($tdTom^-/Iba\text{-}1^+/TMEM119^-$, blue open arrowhead). Scale bar = 20 µm. (d) Plot of host macrophage density at control (Ctrl), 1 day (1d), 1, 2, 4, 6, 12 and 24 weeks post-HSCT (mean ± SEM; two-sided Kruskal-Wallis with Dunn's multiple comparisons test; $^{**}P < 0.01$; P = 0.0063 12 weeks vs. Ctrl; n = 7, 7, 4, 4, 4, 6, 6 and 6 mice for Ctrl, 1d, 1, 2, 4, 6, 12 and 24 weeks, respectively). (e) Plot of donor macrophage density vs. host microglia density at 12 and 24 weeks post-HSCT. (Mean of individual brain sections; simple linear regression [black and orange dashed lines] significance from slope = 0, P = 0.5382 F = 0.3930 $R^2 = 0.02026$ and P = 0.0007 F = 13.86 $R^2 = 0.2958$ for 12 and 24 weeks, respectively; n = 24 and 35 brain sections for 12 and 24 weeks, respectively). (f) Macrophage convex hull plot at 12 and 24 weeks post-HSCT (mean ± SEM; two-sided unpaired t-test; n = 5 and 6 mice for 12 and 24 weeks post-HSCT, respectively). (g) Convex hull plot of donor and host parenchymal macrophages (mean ± SEM; two-sided unpaired t-test; n = 1, 5 and 6 mice for 6, 12 and 24 week donor cells post-HSCT, respectively [binned], and n = 4 and 6 mice for 12 and 24 week host cells post-HSCT, respectively [binned]).



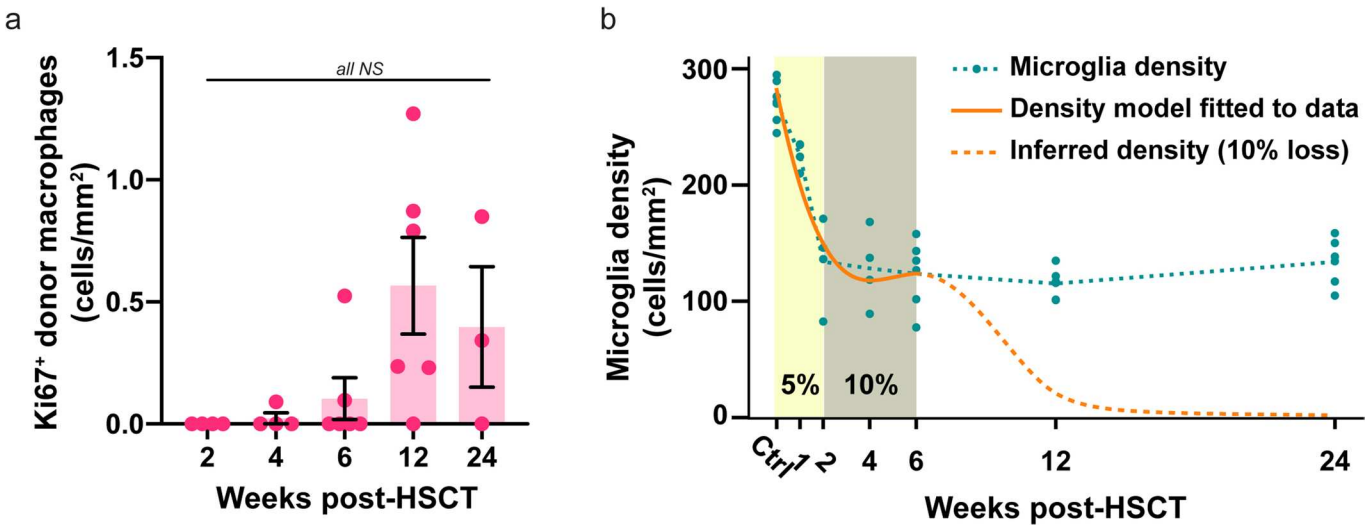
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Effect of HSCT on SVZ proliferation, adult neurogenesis neural progenitors and derived immature neurons. (a) lateral ventricle (LV) sub-ventricular zone with immunohistochemistry for Ki67$^+$ cells (open arrowhead). Scale bar = 100 μm. (b) Plot of Ki67$^+$ density in subventricular zone of the lateral ventricles (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; ****P < 0.0001 Ctrl vs. all; n = 4, 3, 4, 4, 4, 6, 6 and 3 mice for Ctrl, 1d, 1, 2, 4, 6, 12 and 24 weeks, respectively). (c) Adult hippocampus dentate gyrus region showing doublecortin (DCX) positive adult neurogenesis neural progenitor cells/immature neurons in the subgranule zone (open arrowhead). Scale bar = 100 μm. (d) Adult olfactory bulb in control (Ctrl) doublecortin labeling (open arrowhead). Scale bar = 200 μm. (e) Plot of doublecortin adult neurogenesis neural progenitor cell density in dentate gyrus subgranule zone of hippocampus with orange circles representing modeled normal age-related decline of adult neurogenesis DCX. (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; **** p < 0.0001 Ctrl vs. 6, 12 and 24 weeks; n = 4, 3, 4, 3 mice for Ctrl, 6, 12 and 24 weeks, respectively). (f) Plot of doublecortin relative fluorescent intensity (% max control) in sub-ventricular zone of lateral ventricles (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; **** p < 0.0001 Ctrl vs. 6, 12 and 24 weeks; n = 4, 3, 4, 3 mice for Ctrl, 6, 12 and 24 weeks, respectively). (g) Plot of doublecortin relative fluorescent intensity ($\%_{max}$ control) in olfactory bulb (mean ± SEM; two-sided unpaired t-test; ***P = 0.0001 Ctrl vs. 24 weeks; n = 4, 3 mice for Ctrl and 24 weeks, respectively).

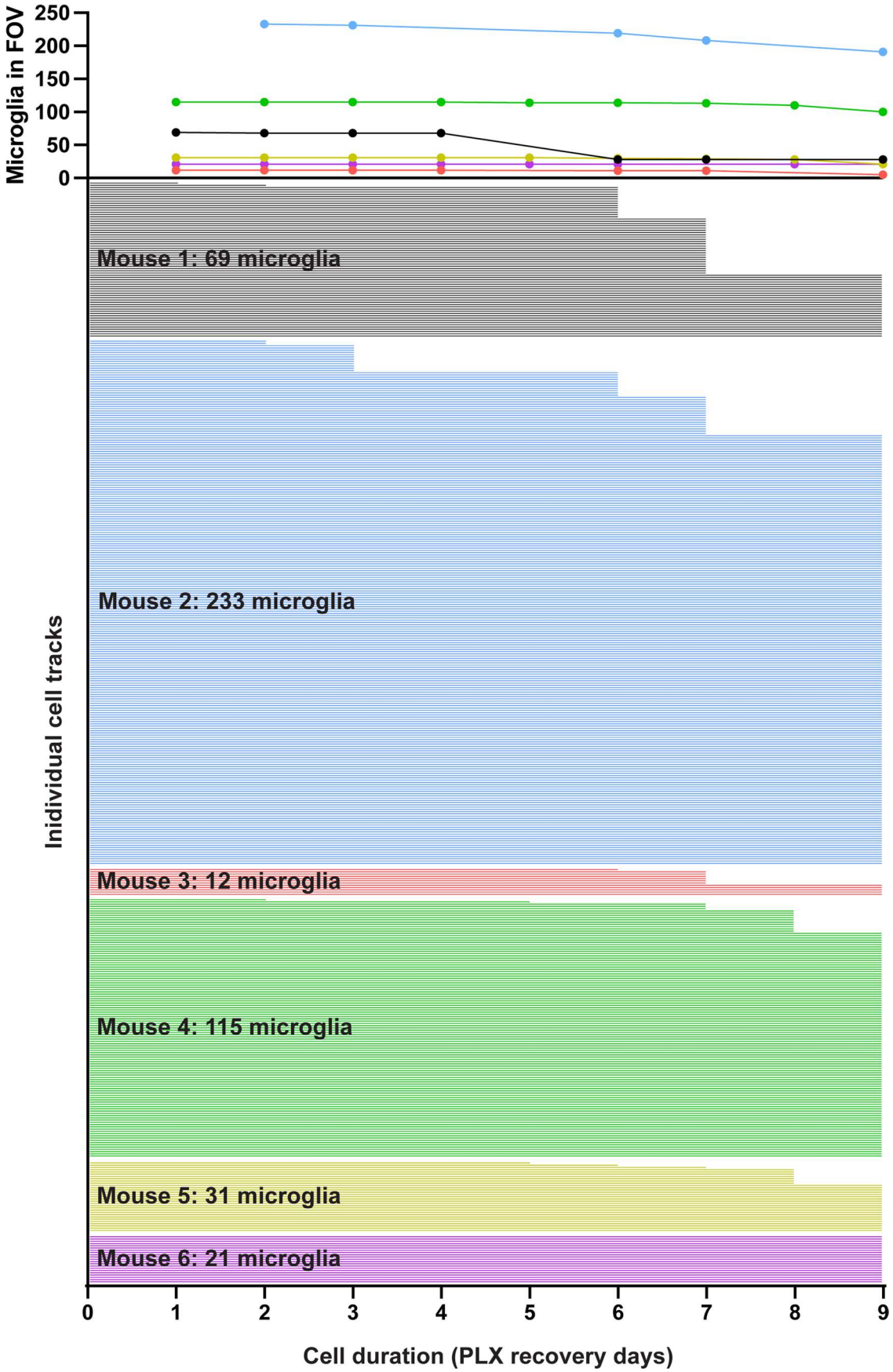


Extended Data Fig. 4 | See next page for caption.

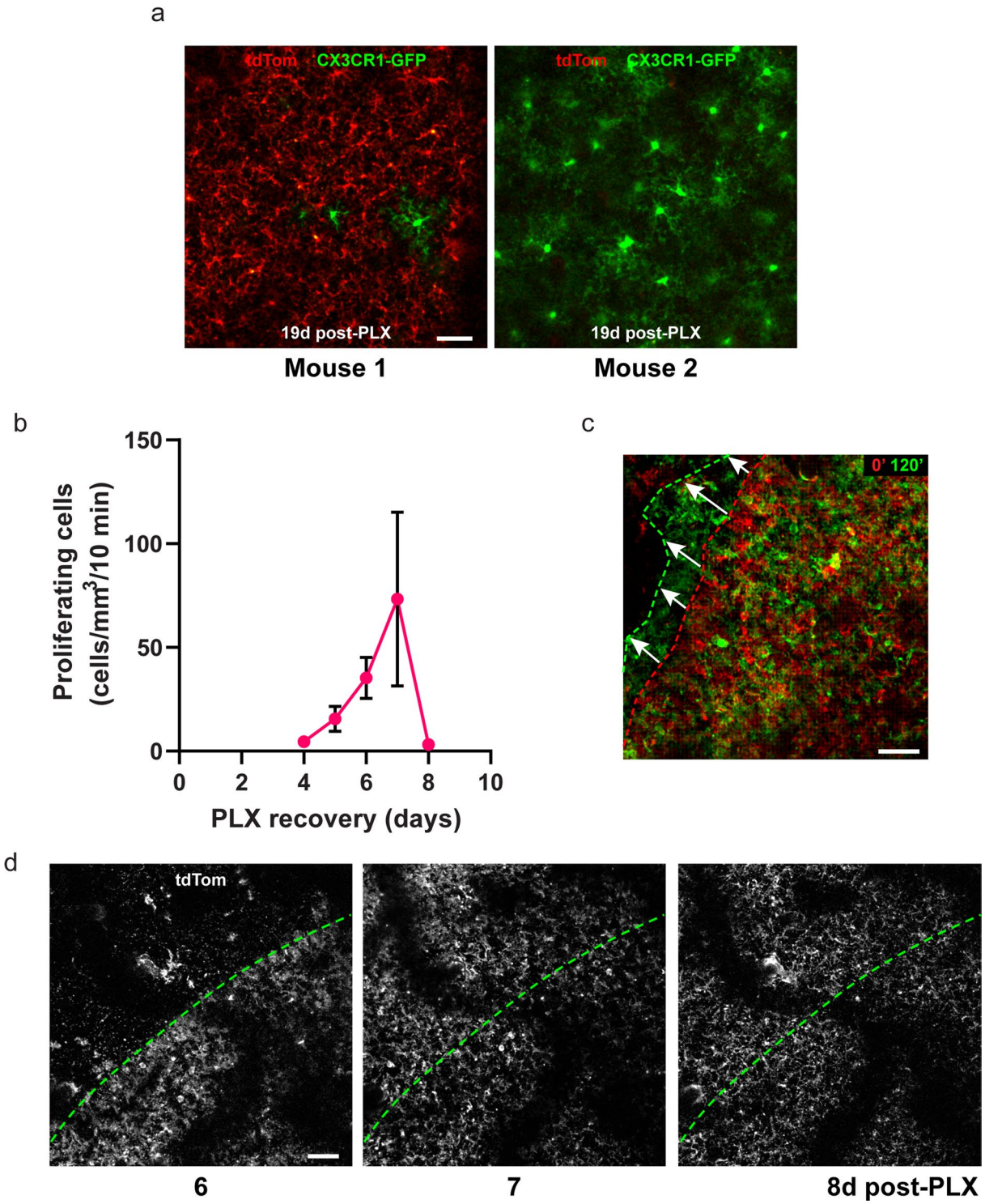
Extended Data Fig. 4 | Ventral to dorsal brain pattern of NG2 depletion and partial recovery with HSCT. (a) Confocal images of NG2 immunohistochemistry of dorsal cortex in control (Ctrl), 4 and 24 weeks post-HSCT with $NG2^+$ cells indicated (open arrowhead). Scale bar = 50 μm. (b) Plot of $NG2^+$ cell density at control (Ctrl), 1 day, 1, 2, 4, 6, 12 and 24 weeks post-HSCT (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; ****$P < 0.0001$ for 1 to 12 weeks vs. Ctrl and **$P = 0.0021$ for 24 weeks vs. Ctrl; n = 7, 3, 4, 4, 4, 6, 6, and 3 mice for Ctrl, 1d, 1, 2, 4, 6, 12, 24 weeks post-HSCT, respectively). (c) Tiled brain hemispheres at Control, 2, 4, 6, 12 and 24 weeks post-HSCT of NG2 cells. Control brain section has dense labeling of NG2 cells throughout cortex. Two weeks post-HSCT most NG2 cells are lost in cortex but small population of dense labeled NG2 cells found in thalamic regions (open arrowhead). By four weeks the section is devoid of NG2 cells. For 6, 12 and 24 weeks the NG2 cells returned but with a ventral reconstitution and a clear demarcation (dashed black line) from dorsal cortical regions. This pattern was observed in most mice, with a few individual outliers having dorsal near-complete recovery (Extended Data Fig. 4b at 6-12 weeks). Scale bar = 1 mm.



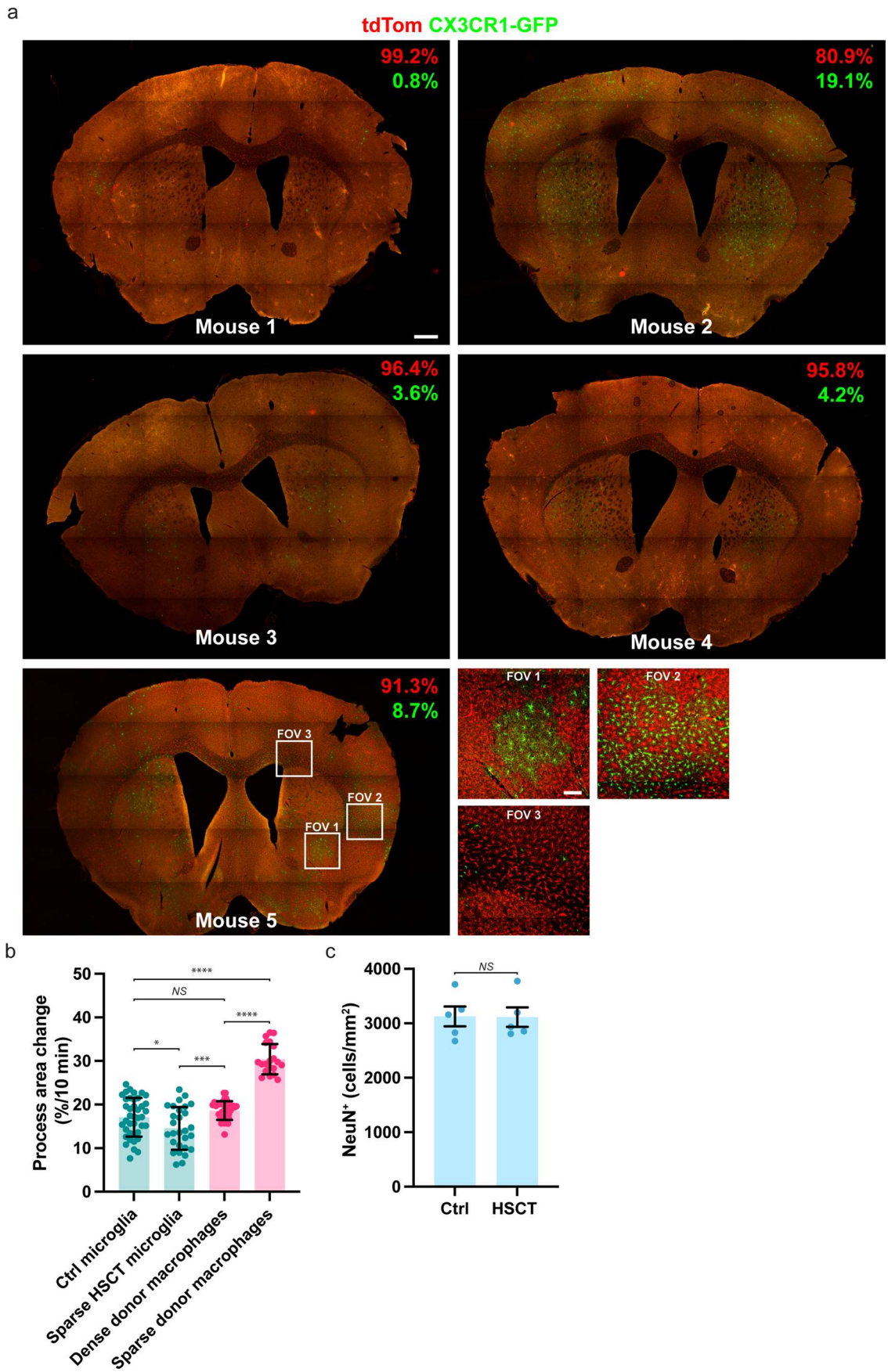
Extended Data Fig. 5 | Effect of HSCT on donor cell proliferation and host microglia Ki67 modeling. (a) Plot of donor macrophage proliferation at 2, 4, 6, 12 and 24 weeks post-HSCT (mean ± SEM; one-way two-sided ANOVA with Tukey multiple comparisons test; n = 4, 4, 6, 6, and 3 for 2, 4, 6, 12, and 24 weeks post-HSCT, respectively). (b) Plot of microglia density (green dashed line and spots; from Fig. 2a) with density model (orange line) at a 5%/day loss rate from Ctrl to 2 weeks post-HSCT and 10%/day from 2-4 weeks post-HSCT as an adjustment for inferred Ki67 increased proliferation. Using this 10%/day loss rate, there would be a complete depletion of host microglia by 12 weeks (orange dotted-line) due to loss of assumed Ki67 driven cell proliferation.



Extended Data Fig. 6 | Single cell tracking of *in vivo* 2-photon image microglia. Scatter plot above of total number of microglia in a FOV for each mouse, with color matched to tracks below, during PLX recovery period. Bottom plot of individual cell tracks (horizontal colored lines) for individual mice from HSCT group for 9 days of PLX3397 recovery. Each line represents one tracked cell and its survival duration.



Extended Data Fig. 7 | Donor cell re-population of brain with effective loss of host microglia, proliferation of donor cells and migration front. (a) *In vivo* 2-photon projected image of mouse 1 with successful donor cell engraftment after PLX-recovery and mouse 2, unsuccessful donor cell engraftment after PLX-recovery showing higher density of host microglia. Scale bar = 50 µm. (b) *In vivo* detected mitotic spheres of proliferating cells segregated into days after PLX3397 recovery from Fig. 4f (mean ± SEM; n = 5 mice). (c) Migration front at 0 min (red dashed line) and 120 min (green dashed line) at 6d post-PLX3397 showing migration direction (white arrows). Scale bar = 50 µm. (d) Same FOV (green dashed line) as in c, at 6, 7 and 8 days post-PLX. Scale bar = 100 µm.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | PLX/busulfan efficient replacement of brain microglia population with donor macrophages and effect on donor and host motility and cortical neuron density. (a) Tiled images of histological sections at 21 weeks post-HSCT of mice from PLX recovery experiment (Fig. 4) showing pattern of tdTom (red) and CX3CR1-GFP (green) cells having an average of 91.27 ± 3.67 % replacement of brain area with donor macrophages. Mouse 5 FOVs highlight typical morphology of host microglia (FOV 1) with large cell area and complex, fine process pattern compared to host peripheral macrophages (FOV 2) having a smaller cell area with less complex process pattern. FOV 3 demonstrates efficient white matter engraftment of donor macrophages. Scale bar for whole-brain images = 500 μm and for FOV images scale bar = 20 μm. (b) Plot of process area change (%/10 min) between control, no HSCT microglia (Ctrl microglia), sparse post-HSCT host microglia (sparse HSCT microglia), post-HSCT/post-PLX macrophages (dense donor macrophages) and post-HSCT macrophages (sparse donor macrophages). (mean ± SD; one-way two-sided ANOVA with Tukey multiple comparisons test; *P = 0.0408 Ctrl microglia vs. sparse HSCT microglia, ***P = 0.0002 sparse HSCT microglia vs. dense donor macrophages, and ****P < 0.0001 for dense donor macrophages and control microglia vs. sparse donor macrophages; n = 40, 28, 40 and 20 cells for Ctrl microglia, sparse HSCT microglia, dense donor macrophages and sparse donor macrophages, respectively). (c) Plot of $NeuN^+$ cell density of layer 2/3 in dorsal cortex between control (Ctrl) and 21 weeks post-HSCT, PLX withdrawal donor macrophage enriched (HSCT) brains. (mean ± SEM; two-sided unpaired t-test; p = 0.9780).

nature research

Corresponding author(s): Kurt A. Sailor, Pierre-Marie Lledo, Nathalie Cartier

Last updated by author(s): December 20, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Zeiss Zen Black software on a Zeiss LSM 207 confocal microscope was used for histological imaging and saved as .lsm files for analysis. PrairieView 5.4.64.100 was used for 2-photon imaging on a Prairie (Bruker) Investigator microscope with raw acquired images converted to an uncompressed .tif format for image analysis.
Data analysis	Imaris 9.1.2 software was used for 3D analysis of in vivo imaged datasets. The spot analysis tool was used to perform a fine registration and then used for tracking individual cells across multiple timepoints within the 3D volume. NVidia 3D vision stereo-glasses were used to provide depth perception. FIJI ImageJ 1.53 was used for histological analysis using the polygon selection tool to select and measure the area while clearing the area outside the selection. The "synchronize windows" plugin and "Cell Counter" plugin were used to quantify the multiple channels for colocalization IHC analysis and the polygon tool was used to trace the microglia projection area. Matlab 2018a was used for image normalization of in vivo image stacks normalizing image intensity along all z-slices and timepoints within a volume (Song & Sailor et al. J.Neurosci, 2013). Slicer 4.10.2 was used for image registration of in vivo imaged volumes prior to Imaris analysis. Prism version 9 was used for data plotting and statistical analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data used in the figures for this study are included in Extended Data Table 1 and are covered by the Creative Commons Attribution 4.0 International License (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical tests were utilized for determining sample size. The size of the samples was based on previous histological and in vivo analysis of brain tissue. At least 4 replicates of each animal were quantified for histological analysis and computed as the mean. In vivo imaging data were analyzed in the same brain volume for each animal across time points, thus causing less variability. All sample sizes are provided in the figure legends.
Data exclusions	No animals, potential outliers or any data were excluded from this study.
Replication	Whenever possible, we confirmed our histological analysis with in vivo analysis in parallel, for example, with observing the 2 week halving of microglia after HSCT in Fig. 2a histology being confirmed in vivo in Extended Data Fig. 2b and Fig. 5c. For microglia senescence, we used immunohistochemistry to observe the expression senescence markers and in vivo analysis to confirm these cells being senescent showing them incapable of cell division. Histological analysis was performed on at least 4 replicates per animal and convex hull measurements were performed on 5 random cells from each replicate having >20 cells analyzed per animal. In vivo PLX-depletion and recovery was performed with two separate cohorts at two different experimental series of dates. Mice 5 and 6 in Extended Data Fig were in cohort 1 with cohort 2 experiments performed 6 months later with mice 1-4, having results similar between groups with depletion and a complete loss of microglia recovery.
Randomization	The mice in this study were randomized and no inclusion criteria were used to separate between experimental groups. All experiments were also performed on male mice, thus no potential gender variability was induced.
Blinding	Histology quantification was blinded by confocal images being assigned an ID number and stored in a separate database. The quantification was blinded to the group allocation. After quantification, the analysis results were divided into experimental groups based on the image ID numbers. In vivo analysis was acquired in across multiple timepoints for the same volume. As a requirement for image registration and tracking methods, the timepoints could not be randomized or blinded within the same animal.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Primary antibodies used: Iba-1 (1:500, guinea pig, Synaptic Systems, 234004), TMEM119 (1:250, rabbit, Abcam, ab209064), Ki67 (1:250, rat, Invitrogen, 14-5698-80), doublecortin (1:500, guinea pig, Millipore, AB2253), NG2 (1:400, rabbit, Chemicon, AB5320), phospho-gamma-H2A.X (1:1k, rabbit, Abcam, ab2893), p21 (1:250, rabbit, Abcam, Ab188224), MCM-2 (1:250, mouse, BD Biosciences, 610701), PCNA (1:250, rabbit, Abcam, EPR3821), pHH3 (1:500, rabbit, Invitrogen, PA5-104936), lamin B1 (1:500, rabbit, Abcam, AB16048) and NeuN (1:500, mouse, Millipore, MAB377). Secondary antibodies used: donkey anti-guinea pig Alexa Fluor 648 (1:500, Jackson Immno Research, 706-605-148), goat anti-rabbit Alexa Fluor 568 (1:500, Invitrogen, A11036), goat anti-rabbit Alexa Fluor 488 (1:500, Invitrogen, A11034), goat anti-rabbit Alexa Fluor 647 (1:500, Invitrogen, A21244), goat anti-rat Alexa Fluor 568 (1:500, Invitrogen, A11077), goat anti-mouse Alexa Fluor 586 (1:500, Invitrogen, A11031).
Validation	The antibodies were validated based on the staining pattern using the manufacturer's recommended dilutions and previously published studies: Iba-1 (Hüttenrauch et al., Acta Neuropathol Commun. 2018 Oct 19;6(1):108), TMEM119 (Bennett ML et al., PNAS. 2016 Mar 22;113(12):E1738-46.), Ki67 (Sobecki M et al., Elife. 2016 Mar 7;5:e13722.), doublecortin (Cho KO et al., Nat Commun. 2015 Mar 26;6:6606.), NG2 (Gautier HO et al., Nat Commun. 2015 Oct 6;6:8518.), p21 (Riessland M et al., Cell Stem Cell. 2019 Oct 3;25(4):514-530.e8.), MCM-2 (Berg DA et al., Cell. 2019 Apr 18;177(3):654-668.e15.), PCNA (Tominaga T et al, PLoS One. 2019 Mar 11;14(3):e0213673.), pHH3 (Park CH et al., PLoS One. 2012;7(9):e44307.), Lamin B1 (Takamori Y et al., IBRO Rep. 2018 Nov 5;5:99-109.) and NeuN (Ge S et al., Nature. 2006 Feb 2;439(7076):589-93.). For the chromagen labeling kit: SA-beta-galactosidase (Tominaga T et al., PLoS One. 2019 Mar 11;14(3):e0213673.).

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mice, all male, >10 weeks of age, Recipients: C57BL/6J and CX3CR1-GFP (B6.129P2(Cg)-Cx3cr1GFP, Jackson Laboratory stock #005582), Donor: mTmG (B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo, Jackson Laboratory stock #007676). The mice were housed in standard cages with a 14 hour light and 10 hour dark cycle at ambient temperature (21 C) and humidity (40-60%).
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used.
Ethics oversight	All experimental procedures followed approved guidelines regulated by the institutional and national committees (Comité d'éthique en expérimentation animale #89).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Part 2: A new approach to microscopy analysis at the frontier of Virtual Reality and human-in-the-loop processing

The linking of learning dynamics to the evolution of neural trees is hampered by the difficulty and time-consuming nature of microscopy image segmentation. Rapid advances in machine learning hold out hope for automatic segmentation[1,2] for many microscopy modalities. However, the need for perfectly standardized images, with identical structures, and above all in large quantities, is a hindrance in neuroscience[3–5]. In particular, there are no databases containing the thousands of structures that would allow using a classic supervised machine learning approach.

We used the ilastik[6,7] software as inspiration for a one-shot learning approach with limited annotations. This software is not easy to use for 3D data, especially if the structures have geometric complexity or significantly varying density profiles in space. Thus, we decided to exploit recent results from the laboratory in the field of virtual reality to annotate the data and make one-shot segmentations. We created a new DIVA-cloud[8] software library that combines the virtual reality implemented in the DIVA platform[9], a python package allowing interaction in virtual reality while viewing data on the cloud, and the ability to perform one-shot learning on this same cloud. This approach combines a native 3D approach, allowing experienced users to quickly annotate structures of interest, and one-shot learning[10,11] to extend these annotations to the entire sample. The model is refined iteratively by the user correcting the automated annotations and the refined model rerun to segment the images.

1 Introduction: Virtual Reality for Everyone

Virtual reality (VR) is by no means a new technology. It has played a role in a number of specialized industrial applications since the 1970s, most notably in the automotive and military sectors. However, recent major shifts in technology have made it accessible to new areas of application such as education, psychiatry and scientific research. It is the convergence of three trends that has made this possible. First, desktop computers now have graphics cards powerful enough to meet the real-time rendering requirements of VR. Secondly, the latest iteration of VR technology has been optimized to ensure visual comfort and usability, minimizing the motion sickness associated with VR use. Finally, the widespread availability of VR headsets such as the HTC Vive, Oculus Rift and Windows Mixed Reality at affordable prices has democratized its use. In a nutshell, VR is a technology that uses visual perception of the real world in fully artificial computer-generated environments. It is a combination of 3 effects: i) An immersion experience. Users wear an optically sealed helmet designed to block out extraneous light. ii) Stereoscopic vision. Each eye sees the same scene from a slightly different angle, mimicking the way our eyes see the world in three dimensions. iii) Motion capture. The positions of the user's head and controllers are detected with three or six degrees of freedom, allowing accurate tracking of movements in rendered scenes.

Interaction in VR environments is not standardized, and significant work is still required to establish efficient modes of interaction. The use of VR controllers is critical in this regard. It provides a means to perform basic tasks such as grabbing, scaling and rotation, although data-specific challenges still exist (*e.g.* changing display parameters of the transfer function). As the use of VR technologies will increase for scientific research, automating some procedures for visualization and navigation will become appealing and, in some cases, necessary. Several publications have already investigated the reaction of users using VR for specific tasks[14,15]. In our experience, we have observed consistent trends in user behavior when trying our VR applications. For example, it is not uncommon to witness users standing up, moving the data vertically, moving around the data in order to explore its spatial extents before manipulating it. Automated protocols to encourage initial data exploration can aid in this regard.

Through these effects, VR provides a unique viewing and interacting context that enhances the perception of volumetric detail. Compared to visualization on a conventional 2D screen, where users typically view three-dimensional (3D) data as passive observers, the immersive VR experience allows users to literally enter their data in a way that feels spatially realistic (see Fig. 15).

2 Data Visualization

The most direct advantage of VR for visualizing scientific data is the freedom it grants users to intuitively explore and interact with their environment. Objects can be observed from arbitrary vantage points, as in the real world. Interaction using specialized VR controllers available with modern headsets can be performed with millimeter precision. This sense of realism translates to a fluid and natural-feeling experience for the user, in turn allowing otherwise complicated spatial tasks to be performed rapidly, often magnitude faster than with conventional computer and monitor configurations. Moreover, moving within a VR scene increases the likelihood to detect patterns of interest. VR naturally encourages exploration and curiosity-driven action.

The generation of a VR scene comes at a high computational cost. A different scene has to be rendered for each eye (*i.e.* to allow stereoscopy) at a refresh rate of at least 60 Hz (often ~80 Hz is recommended) for smooth navigation. The most common approach for ameliorating rendering speed is by reducing the image quality, although graphics card manufacturers such as NVIDIA propose VR-specific features such as single-pass stereo rendering, headset lens-matched pixel resolution and multiple graphics card configurations to improve rendering performance. The visualization of volumetric image stacks presents some particularly difficult challenges, especially when done without pre-treatment or segmentation. The first of these is the design of a 3D look-up table or *transfer function* for efficient data visualization. The transfer function assigns optical properties such as color and transparency to each voxel of the visualized scene. It defines the physical rules of light propagation in the sample.

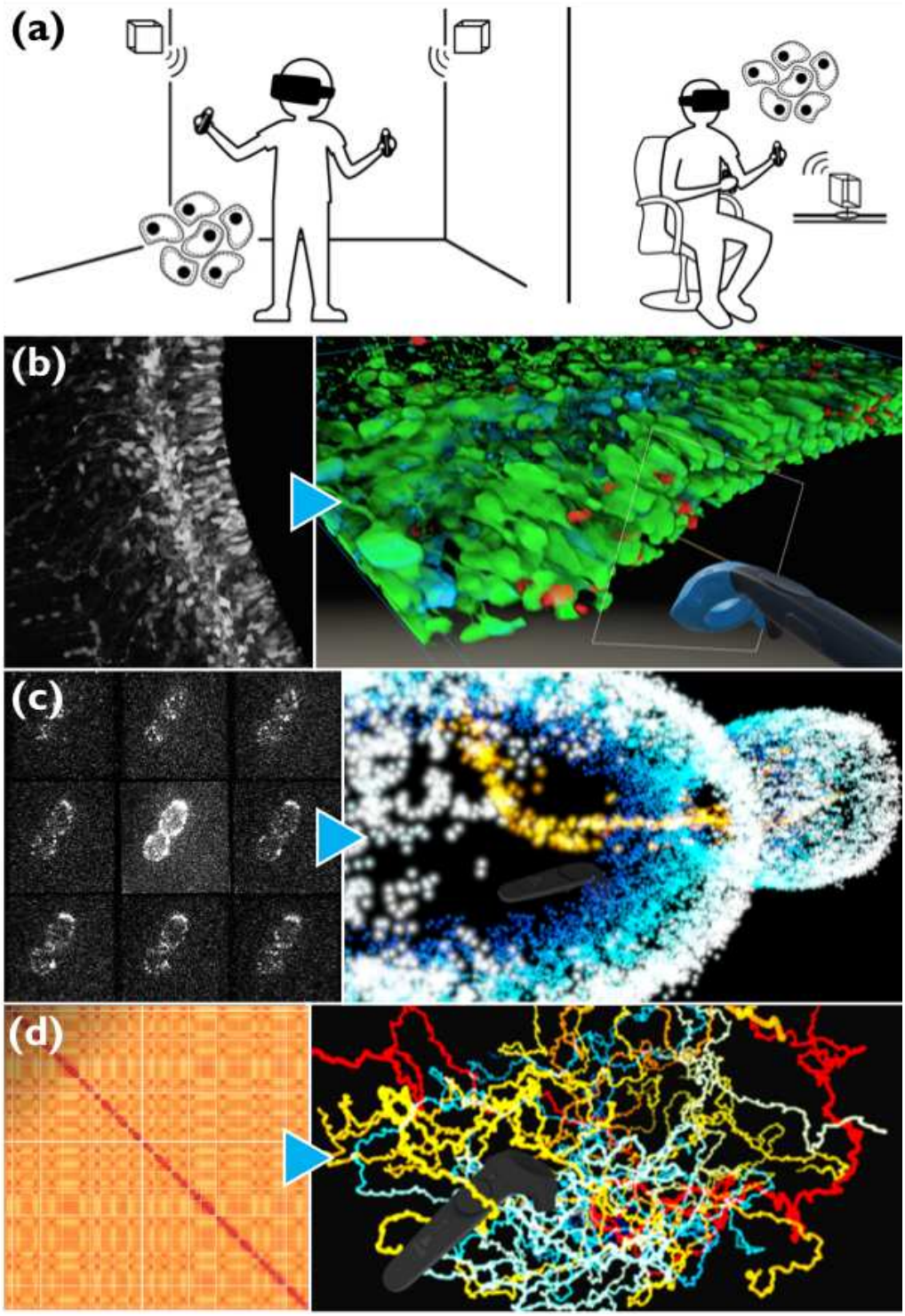


Figure 15: Visualizing data in VR taken from Ref[17]. (a) Schematic illustration of usual usage configurations for VR. The user can either be in a standing (room-scale) (left) or sitting (right) posture. The user wears a stereoscopic headset that permits total immersion in a computer-generated environment. External infrared sensors and internal sensors ensure precise rotational and translational motion tracking. The user visualizes

data, here a set of cells in 3D, and can interact with the representation using one or two VR controllers. Experience showed us to favor one VR controller for ease of use. (b) Example visual representations of microscopy data. On the left, a maximum intensity projection of an image stack of neurons imaged by spinning disk confocal microscopy[12]. On the right a capture of the same dataset in VR. There is a significant increase in volumetric cues and information in the VR mode; 3D geometry is accessible, boundaries can be enhanced and complex geometrical relations between neurons are clearly observed. (c) Single-molecule point cloud originating from multifocus microscopy[18]. On the left, the raw image acquisition. On the right, the VR representation. In blue-white, the wall of a Saccharomyces cerevisiae yeast cell, in red-yellow a filament of alpha tubulin passing between the dividing mother and daughter cells. (d) Volumetric representation of the simulated positions of sixteen saccharomyces cerevisiae yeast chromosomes (distinguished by color) based on the statistical properties in a Hi-C map, an example of which is shown on the left[13] All VR screen captures were generated with the DIVA software platform.

Moreover, it is crucial to ensure optimal information representation in the visualization of the 3D image. This can mean allowing structures inside a cell to be seen from the outside while revealing the barriers between domains of different properties. An example of such a challenge is shown in a complex microscopy image of the mouse neocortex[12] (Fig. 15). Immersive visualization can also facilitate complicated segmentation tasks of microscopic structures within image stacks. In Fig. 16, we show a segmented portion of neurons in the mouse olfactory bulb imaged using confocal microscopy. Here, surfaces have been pre-annotated by the user and the transfer function has been optimized to ensure their visibility. Combined with stereoscopy, a simple glance reveals both geometric and some topological features of this local network. In Fig. 16, we also show an example of a user visualizing and interacting with a macromolecule in VR. Similarly, we show 3D yeast chromosome geometry based on statistical properties of Hi-C maps[13] in Fig. 15. Besides visualization, navigating in the data is a second major challenge. Figures 15 and 16 display screen captures of images rendered in the VR headset.

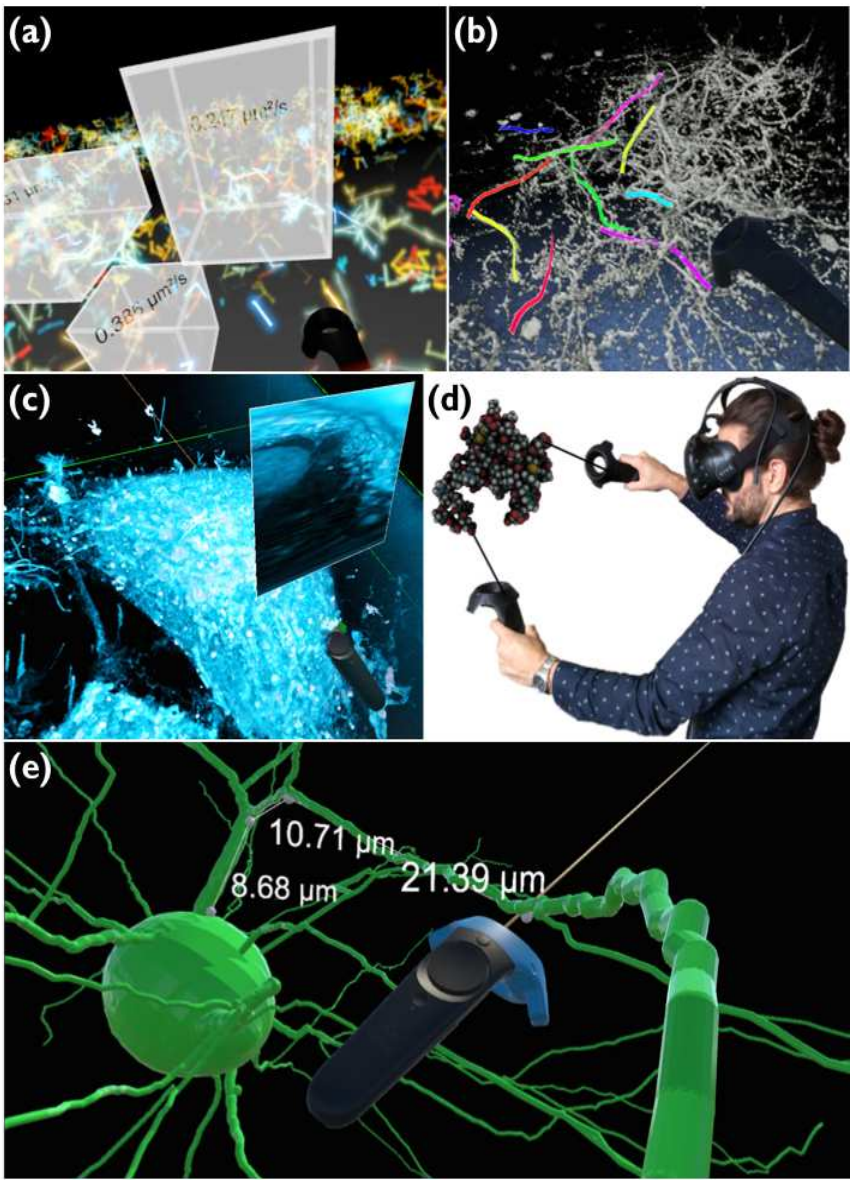


Figure 16: Data treatment in VR: a human-in-the-loop approach taken from ref[17]. (a) Bayesian inference of local diffusivity of tracer fluorescent beads in the nucleus of a U2OS cell using VR. A video demonstration can be seen at http://goo.gl/dnNueu. The user identifies a local region of interest (ROI) in 3D, uses the VR controller to define the limits of the ROI and triggers a program (based on the InferenceMAP software[16]) that computes the diffusion using a Bayesian inference approach. The results are then overlaid. (b) UnityMol VR[20]. This software allows the user to visualize and interact with large molecules in VR. (c) Use of VR in a human-in-the-loop segmentation task. Raw confocal images of fluorescent mouse olfactory bulb neurons with a reduced signal-to-noise ratio. VR allows easy determination of axon position and direction permitting a rapid segmentation with the VR controller. (d) Projection tool aligned to the VR controller used to map the inside of a plated HeLa cell in a serial block-face scanning electron microscopy image stack[28]. (e) Precise measurement of axon segments using the VR controller in a segmented mouse neuron database. All VR screen captures were generated with the DIVA software platform.

3 Data treatment and numerical simulations

The visual intuition provided by VR is not limited to facilitating spatial understanding of complex volumetric data. It provides a canvas for data processing and interaction. It is at the frontier of human cognition, data visualization and direct processing that VR could have its greatest impact. Indeed, it is in research that many applications seem promising for VR. The absence of large volumes of initial data and the need to actively explore these data in order to understand the important elements push to a mixture of exploration of these in a 3D environment and human intervention to treat the data.

A simple example of direct interaction is the selection of 3D regions of interest (ROIs) for export and analysis. 3D imaging of single molecules is an efficient method to probe complex biological processes at the nanoscale in cellular environments. The identification of 3D patterns (static or dynamic) indicative of biological activity is not an easily automatable task. Figure 16 illustrates an example where the user defines ROIs using the VR controller, on which an inference program is run to estimate the local diffusivity value of all single molecules moving in that volume[16]. Similarly, being immersed in the data provides valuable insight into data with complex geometry. In Fig. 16, we show an example of a user efficiently mapping the interior of a mammalian cell using the VR controller to perform a planar projection into a high-resolution electron microscopy dataset. By being able to arbitrarily position the projection plane with the VR controller, this function allows structures of interest to be easily revealed to the user. Finally, and similarly to the previous example, the user can directly create 3D annotations of data and perform simple measurements. We show an example in Fig. 16, where dendrites of a segmented neuron are interactively and precisely measured. These annotations can be exported and fed to other programs to characterize the geometrical properties of the tagged structures. Notably, all these examples demonstrate that we can transform intuition and perception into quantitative results in a matter of seconds.

Another possibility is to couple human interaction with simulations to understand the processes driving the dynamics of living systems. A direct example is molecular dynamics (MD). The high-dimensional energy landscapes associated with biomolecule

folding problems entail a significant combinatorial complexity of obtaining solutions. Optimized MD software packages show continuous improvements, but MD remains a computationally intensive task, especially for large molecules[19,20]. Introducing visual intuition and external knowledge can thus massively improve performance. In Ref[21] the authors created a VR software platform where interactive simulations enable multiple users to simultaneously manipulate molecules and observe their dynamics, computed in the cloud from physical equations of motion. Here, VR provides geometric and organizational intuition, wherein the user can pull, push, introduce or release molecules and observe the resultant physical effects of the interactions. Note that the authors added an innovative "multi-users presence" in VR as a component of their software. This way researchers can learn and interact in virtual environments together while they manipulate the molecules. Similarly, UnityMol[20] is a software package that brings VR and human interaction to molecule editing and prototyping, a photo of which is shown in Fig. 16.

Finally, an important concept is *human-in-the-loop* (HITL) data treatment. HITL means that the user and the machine (*i.e.* learning algorithm) work simultaneously to achieve a specific data treatment task[22–24]. We predict that such mixed approaches will be instrumental in the context of explainable artificial intelligence (xAI), where human intuition and reasoning are used to justify predictions in artificial intelligence systems. Motivation for HITL approaches derives from two often-encountered scenarios in research. In the first, data is complex to acquire, non-standardized and limited in number. Developing fully automated algorithms or leveraging machine learning in these situations is challenging. The time required to develop efficient algorithms can exceed the time for semi-manual analysis in its own right. Furthermore, generalization to new data is not ensured due to complexities in data acquisition protocols and the limited number of examples to train on. Relating physical interactions to learning algorithms in VR may greatly reduce the time to treat data, improve the efficiency in performing tasks and reveal possible strategies to automate the treatment. For example, in ref[25] the authors developed a manual interface to trace neurons from 3D image stacks. VR allows for easy visualization of neurons in 3D, even within noisy confocal microscopy datasets. In this context, the user traces neurons by simply moving the VR controller within the data as if it were a real object. Eventually, human intervention can be used to simply define initial

“starting points'' for the segmentation task, after which an algorithm will perform the optimization. Human intervention can also be intermittent in the training of a learning algorithm, similar to the workflow described in software such as ilastik[7] and others: VR would be used for fast data tagging in 3D, and afterward the algorithm locally learns rules to classify data.

HITL is equally applicable in cases of massive, complex and standardized data. A direct example is the collaborative efforts to trace neurons from EM serial imaging[26,27]. As mentioned above, VR human interaction will most likely be used to accelerate data tagging. Interacting and learning is possible on small portions of data loaded in the VR environment, which can then be propagated to the remainder of the dataset. Several challenges will still have to be resolved, namely loading and visualizing image files that are significantly larger than the live memory of the graphics card. Furthermore, it is challenging to ensure the efficiency and robustness of results across multiple users performing tagging, testing and visual verification.

Another important feature of treating data in VR is the transfer of expert knowledge to uninitiated users. Immersive environments engage the user to interact with the data. Active interaction promotes better learning than passive observation. The possibility of following the actions of experts performing tasks while being inside the data can be transformative in training and education. Mutual interaction in the environment can also drastically reduce learning curves.

4 DIVA: Data integration and visualization in augmented and virtual environment

We introduce here briefly the underlying principle of the DIVA platform. DIVA is a user-friendly software platform that automatically creates detailed three-dimensional reconstructions of raw experimental image stacks that are integrated in virtual reality. In DIVA's immersive virtual environment, users can view, manipulate and perform measurements on their image stack as they would real physical objects. The software provides high-quality volume rendering with native TIFF file compatibility. DIVA does not

require pretreatment or segmentation of the data to provide representation, nor does it require conversion to intermediate file formats, as is the case for most 3D image visualization software. Accordingly, it can be used to visualize any type of microscopy image regardless of the imaging modality, ranging from confocal to EM images and any type of medical imaging data from CT-scans to MRIs.

The software architecture of DIVA, entitled *Lean Mappe*r, consists in arranging distinct subsystems to interact with each other via a top-level supervisor. Specifically, a data subsystem, which holds loaded imaging data, is decoupled from a user-facing interface subsystem. A unique data identifier is used as a reference by the supervisor to trigger events. For example, when the user adjusts a slider to increase the ambient lighting in the interface sub-system, the corresponding event is emitted to the data subsystem, which alters the rendered values in real-time. A main advantage of this data-oriented approach is low-latency user interactions with data and fluid navigation.

An image stack loaded in DIVA is rendered as a 3D volume through a bilinear texture interpolation. The user can control the aspect ratio of the volume by specifying the pixel size and stack spacing in a user interface panel.

DIVA is developed using the Unity game engine[30]. DIVA includes a dual interface: a desktop mode to use a standard computer monitor and a VR mode where the user interacts with data in the VR headset. Design decisions were motivated by the realization that, despite continuous improvements of VR technology, spending large amounts of time in VR space is not a comfortable experience. In DIVA, the VR interface is tailored to efficient scientific analysis of imaging data, while the desktop interface focuses on parameter settings and initial visual screening of the data. Hence, features of the software are implemented in each respective mode separately.

The desktop interface allows the user to modify visualization parameters such as voxel size and lighting. It also includes a user-friendly transfer function interface, which allows users to specify pixel intensity-dependent opacity and color by means of absorption and emission curves, respectively, in real-time. In this interface, data appear as a 3D volumetric representation on a 2D screen. The volume can be translated, rotated, and scaled using the mouse.

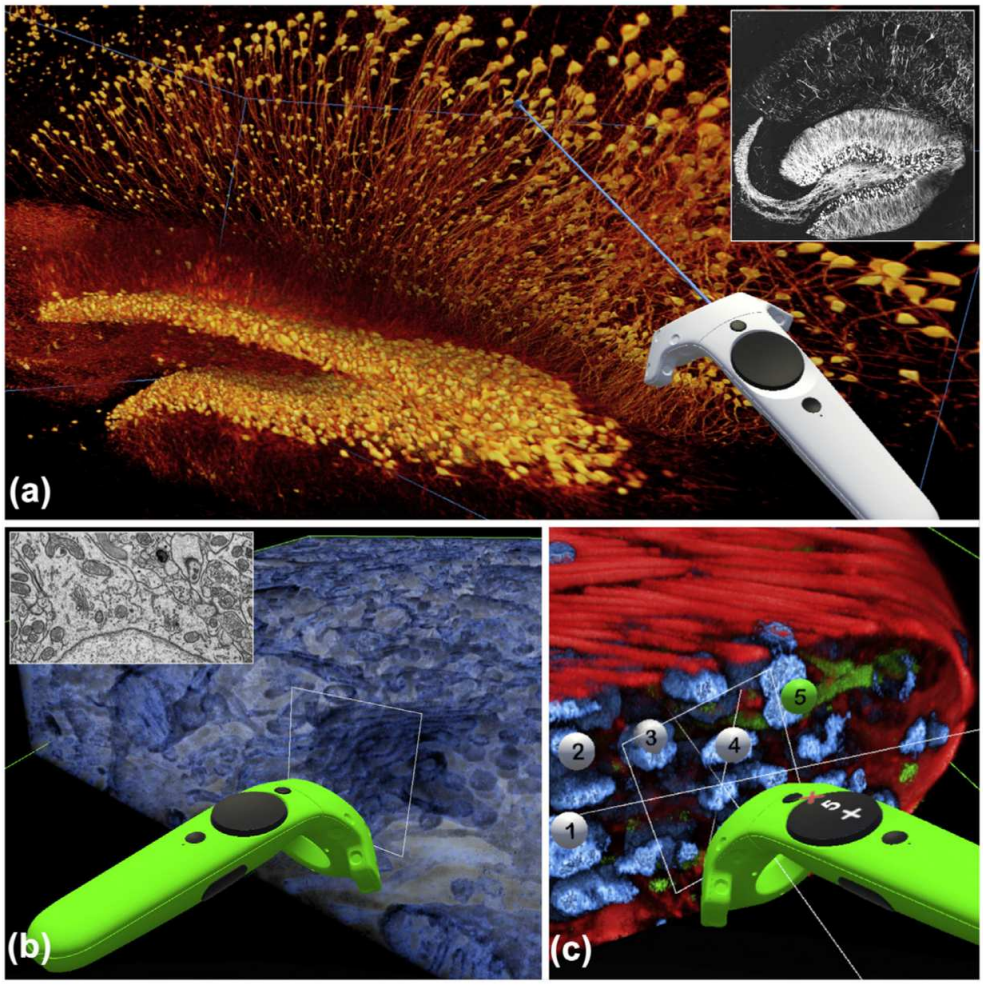


Figure 17: Example of DIVA usage from ref[9]. VR visualization of raw microscopy TIFF image stacks using DIVA. (a) Mouse hippocampus imaged by two-photon confocal microscopy (Thy-1-GFP mouse) with a slice from the raw image shown in inset[29]. (b) Focused ion beam scanning EM of components of an adult mouse neuron; the Golgi apparatus and mitochondria are seen in the image (raw image is shown in inset). (c) Multichannel confocal acquisition of Drosophila testis (red corresponds to actin, blue to nuclei and green to fusome). The positions of nuclei are labeled using a counting tool included with DIVA. In all of the panels, the color of the VR controller is associated with the action being performed.

In the VR interface, DIVA renders image stacks as virtual "physical objects" in a virtual room environment. The user can grasp the image stack and navigate inside it via physical manipulation using the VR controller (bundled with all commercial VR headsets). A key VR function included in DIVA to address the complexity of biological and medical images is a handheld clipping plane that allows interactive removal of portions of the rendered volume in real-time. This feature is essential when extracting information from

dense images. For example, the perception of EM images benefits greatly from DIVA. The flexible transfer function interface allows a smooth modulation of pixel-dependent transparencies and colors to reveal structures of interest while suppressing undesired pixel information such as the significant background noise characteristic of electron microscopy (EM) images. In practice, there is a significant design constraint, as volume rendering techniques are computationally costly and must be performed twice during each frame in VR (*i.e.* once for each eye). To address this challenge, DIVA renders volumes through an optimized and simplified ray-casting approach whose sampling resolution is easily adjusted in real-time to ensure a sufficient frame rate for the user[31]. Additionally, the rendering resolution of the image is dependent on how far the camera (observer) is; the image is better resolved in regions in which the user is interacting or observing. This makes it possible to maintain a high frame rate even on modest computer configurations. The current available version of DIVA supports visualization of up to four different channels, whose color and absorption (transfer function) characteristics can be customized individually.

5 A quick introduction to related works both in microscopy and medical imaging

The access to affordable VR headsets, efficient graphics cards and easily accessible software development platforms[30,32,33] have opened up access to VR developments for a much wider audience. These factors have promoted initiatives combining imaging techniques with VR in order to address topics in cell biology. As a result, VR applications are often predicted to become essential components of the experimental research environment[17,34].

3D Image stack visualization within the VR environment is the aim of numerous initiatives. These include ConfocalVR[35] and Scenery[36], which can generate volumetric reconstructions of microscopy images, following one or multiple steps of pretreatment. Neuroscience, at various scales, is a domain of interest for VR visualization. Large images of entire nervous systems can now be acquired with optical and electron microscopy[37]

(EM). Tracing the complex neuronal structures is essential as there is a link between structure, connectivity and function of neural circuits. Some initiatives already use VR to address this[25,38]. Applications at the frontiers of microscopy and neurosurgery have also been demonstrated[39–41].

The possibility for multiple users acting on the same data within a virtual environment have been tested in these applications[42]. Additionally, some projects address topic-specific challenges in microscopy, such as colocalization of probes recorded in different color channels[43].

Companies developing software for microscopy image analysis are developing VR extensions to their microscopy software. Major advances are found in the following software[44,45], which include an optimized data interaction interface. Other research applications focus on biomolecule structural information visualizations and interactions[46–49].

In medical applications, VR has found multiple applications in surgery-specific topics, notably education related. While VR may be useful for radiology[50,51], radiologists are trained to perform 3D mental reconstructions of medical images, limiting their interest in immersive visualization modalities. Craniofacial trauma education[52], neurosurgical training[53], spinal surgery[54], anatomy education[55], orthopedic surgery training[56–58], and patient education[59] have all been demonstrated in this regard. Furthermore, the first clinical results hint towards an effective usefulness of VR in surgical applications spanning heart disease[60,61], breast cancer[62,63], liver surgery[64,65], pediatric surgery[47,66], and orthopedic surgery[57,67]. Multiple new companies now focus on, or are being created, with VR and surgery at the core of their business, such as immersiveTouch[68] or surgical theather[69].

6 DIVA-cloud and Genuage analysis

In the following, we present two publications that introduce two data processing procedures combining virtual reality, human-in-the-loop processing and various machine learning procedures.

References

1. Arganda-Carreras, I. *et al.* Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* **33**, 2424–2426 (2017).
2. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
3. Soltanian-Zadeh, S., Sahingur, K., Blau, S., Gong, Y. & Farsiu, S. Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. *Proc. Natl. Acad. Sci.* **116**, 8554–8563 (2019).
4. Vogt, N. Neuron segmentation with deep learning. *Nat. Methods* **16**, 460–460 (2019).
5. Abdollahzadeh, A., Belevich, I., Jokitalo, E., Sierra, A. & Tohka, J. DeepACSON automated segmentation of white matter in 3D electron microscopy. *Commun. Biol.* **4**, 1–14 (2021).
6. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
7. Sommer, C., Straehle, C. N., Köthe, U. & Hamprecht, F. A. ilastik: Interactive Learning and Segmentation Toolkit. in *Eighth IEEE International Symposium on Biomedical Imaging (ISBI 2011).Proceedings* 230–233 (2011). doi:10.1109/ISBI.2011.5872394.
8. Guérinot, C. *et al.* New Approach to Accelerated Image Annotation by Leveraging Virtual Reality and Cloud Computing. *Front. Bioinforma.* **1**, (2022).
9. El Beheiry, M. *et al.* DIVA: Natural Navigation Inside 3D Images Using Virtual Reality. *J. Mol. Biol.* **432**, 4745–4749 (2020).
10. Miller, E. G., Matsakis, N. E. & Viola, P. A. Learning from one example through shared densities on transforms. in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)* vol. 1 464–471 vol.1 (2000).
11. Fei-Fei, L., Fergus, R. & Perona, P. One-shot learning of object categories. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 594–611 (2006).
12. Brault, J.-B. *et al.* Comparative Analysis Between Flaviviruses Reveals Specific Neural Stem Cell Tropism for Zika Virus in the Mouse Developing Neocortex. *EBioMedicine* **10**, 71–76 (2016).

13. Schalbetter, S. A. *et al.* SMC complexes differentially compact mitotic chromosomes according to genomic context. *Nat. Cell Biol.* **19**, 1071–1080 (2017).
14. Padmanaban, N., Konrad, R., Stramer, T., Cooper, E. A. & Wetzstein, G. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proc. Natl. Acad. Sci.* **114**, 2183–2188 (2014).
15. Maloca, P. M. *et al.* High-Performance Virtual Reality Volume Rendering of Original Optical Coherence Tomography Point-Cloud Data Enhanced With Real-Time Ray Casting. *Transl. Vis. Sci. Technol.* **7**, 1–11 (2018).
16. Beheiry, M. E., Dahan, M. & Masson, J.-B. InferenceMAP: mapping of single-molecule dynamics with Bayesian inference. *Nat. Methods* **12**, 594–595 (2015).
17. El Beheiry, M. *et al.* Virtual Reality: Beyond Visualization. *J. Mol. Biol.* **431**, 1315–1321 (2019).
18. Hajj, B. *et al.* Whole-cell, multicolor superresolution imaging using volumetric multifocus microscopy. *Proc. Natl. Acad. Sci.* **111**, 17480–17485 (2014).
19. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2012).
20. Doutreligne, S., Cragnolini, T., Pasquali, S., Derreumaux, P. & Baaden, M. UnityMol: Interactive scientific visualization for integrative biology. *2014 IEEE 4th Symp. Large Data Anal. Vis. LDAV* 109–110 (2014) doi:10.1109/LDAV.2014.7013213.
21. Johnston, A. P. R. *et al.* Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Traffic* **4**, 105–110 (2018).
22. Xin, D. *et al.* Accelerating Human-in-the-loop Machine Learning: Challenges and Opportunities. *arXiv* **abs/1804.05892**, (2018).
23. Ren, S., He, K., Girshick, R. B. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **abs/1506.01497**, (2015).
24. Acuna, D., Ling, H., Kar, A. & Fidler, S. Efficient Interactive Annotation of Segmentation Datasets with Polygon-RNN++. *ArXiv180309693 Cs* (2018).
25. Usher, W. *et al.* A Virtual Reality Visualization Tool for Neuron Tracing. *IEEE Trans. Vis. Comput. Graph.* **24**, 994–1003 (2018).
26. Wanner, A. A., Genoud, C., Masudi, T., Siksou, L. & Friedrich, R. W. Dense EM-based reconstruction of the interglomerular projectome in the zebrafish olfactory

bulb. *Nat. Neurosci.* **19**, 816–825 (2016).

27. Funke, J. *et al.* Large Scale Image Segmentation with Structured Loss based Deep Learning for Connectome Reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* (2018) doi:10.1109/TPAMI.2018.2835450.
28. Boassa, D. *et al.* Chlamydia trachomatis, Human HeLa. *Cell Image Libr.* (2017).
29. Sun, G. J. *et al.* Seamless Reconstruction of Intact Adult-Born Neurons by Serial End-Block Imaging Reveals Complex Axonal Guidance and Development in the Adult Hippocampus. *J. Neurosci.* **33**, 11400–11411 (2013).
30. UnityTechnologies. *Unity*. (2021).
31. Engel, K. *et al.* Real-time volume graphics. in *Proceedings of the conference on SIGGRAPH 2004 course notes - GRAPH '04* 29?es (Association for Computing Machinery Press, 2004). doi:10.1145/1103900.1103929.
32. OpenXR - High-performance access to AR and VR —collectively known as XR— platforms and devices. *The Khronos Group* https://www.khronos.org/openxr/ (2016).
33. Unreal Engine | The most powerful real-time 3D creation tool. *Unreal Engine* https://www.unrealengine.com/en-US/.
34. Matthews, D. Virtual-reality applications give science a new dimension. *Nature* **557**, 127–128 (2018).
35. Stefani, C., Lacy-Hulbert, A. & Skillman, T. ConfocalVR: Immersive Visualization for Confocal Microscopy. *J. Mol. Biol.* **430**, 4028–4035 (2018).
36. Gunther, U. *et al.* scenery: Flexible Virtual Reality Visualization on the Java VM. in *2019 IEEE Visualization Conference (VIS)* 1–5 (IEEE, 2019). doi:10.1109/visual.2019.8933605.
37. Eichler, K. *et al.* The complete connectome of a learning and memory centre in an insect brain. *Nature* **548**, 175–182 (2017).
38. Peng, H. *et al.* BigNeuron: Large-scale 3D Neuron Reconstruction from Optical Microscopy Images. *Neuron* **87**, 252–256 (2015).
39. Mauro, A. de, Raczkowsky, J., Halatsch, M. E. & Worn, H. Virtual Reality Training Embedded in Neurosurgical Microscope. in *2009 IEEE Virtual Reality Conference* 233–234 (IEEE, 2009). doi:10.1109/vr.2009.4811031.
40. Wang, Y. *et al.* TeraVR empowers precise reconstruction of complete 3-D neuronal morphology in the whole brain. *Nat. Commun.* **10**, (2019).

41. Wisotzky, E. L. *et al.* Interactive and Multimodal-based Augmented Reality for Remote Assistance using a Digital Surgical Microscope. in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* 1477–1484 (IEEE, 2019). doi:10.1109/vr.2019.8797682.
42. Naviscope. *Image-guided Navigation and VIsualization of large data sets in live cell imaging and microSCOPy*. (2021).
43. Theart, R. P., Loos, B. & Niesler, T. R. Virtual reality assisted microscopy data visualization and colocalization analysis. *BMC Bioinformatics* **18**, 1–16 (2016).
44. Dekkers, J. F. *et al.* High-resolution 3D imaging of fixed and cleared organoids. *Nat. Protoc.* **14**, 1756–1771 (2019).
45. syGlass. *Scientific Virtual Reality*. (2021).
46. Martinez, X. *et al.* Open framework to explore and manipulate molecular structures and trajectories in VR. *Submitted* (2018).
47. Balo, A. R., Wang, M. & Ernst, O. P. Accessible virtual reality of biomolecular structural models using the Autodesk Molecule Viewer. *Nat. Methods* **14**, 1122–1123 (2018).
48. Goddard, T. D. *et al.* Molecular Visualization on the Holodeck. *J. Mol. Biol.* **430**, 3982–3996 (2018).
49. Cassidy, K. C., Šefčík, J., Raghav, Y., Chang, A. & Durrant, J. D. ProteinVR: Web-based molecular visualization in virtual reality. *PLOS Comput. Biol.* **16**, e1007747 (2020).
50. Elsayed, M. *et al.* Virtual and augmented reality: potential applications in radiology. *Acta Radiol.* **61**, 1258–1265 (2020).
51. Uppot, R. N. *et al.* Implementing Virtual and Augmented Reality Tools for Radiology Education and Training, Communication, and Clinical Care. *Radiology* **291**, 570–580 (2019).
52. Bouaoud, J. *et al.* DIVA, a 3D virtual reality platform, improves undergraduate craniofacial trauma education. *J. Stomatol. Oral Maxillofac. Surg.* **122**, 367–371 (2021).
53. Bernardo, A. Virtual Reality and Simulation in Neurosurgical Training. *World Neurosurg.* **106**, 1015–1029 (2017).
54. Pfandler, M., Lazarovici, M., Stefan, P., Wucherer, P. & Weigl, M. Virtual reality-based simulators for spine surgery: a systematic review. *Spine J.* **17**, 1352–1363

(2017).

55. Uruthiralingam, U. & Rea, P. M. Augmented and Virtual Reality in Anatomical Education – A Systematic Review. in *Advances in Experimental Medicine and Biology* 89–101 (Springer International Publishing, 2020). doi:10.1007/978-3-030-37639-0_5.

56. Walbron, P., Thomazeau, H. & Sirveaux, F. , , Virtual reality simulation“ in der Orthopädie und Unfallchirurgie in Frankreich. *Unfallchirurg* **122**, 439–443 (2019).

57. Bartlett, J. D., Lawrence, J. E., Stewart, M. E., Nakano, N. & Khanduja, V. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? *Bone Jt. J.* **100-B**, 559–565 (2018).

58. Yoo, J. S., Patel, D. S., Hrynewycz, N. M., Brundage, T. S. & Singh, K. The utility of virtual reality and augmented reality in spine surgery. *Ann. Transl. Med.* **7**, S171–S171 (2019).

59. Dyer, E., Swartzlander, B. J. & Gugliucci, M. R. Using virtual reality in medical education to teach empathy. *J. Med. Libr. Assoc.* **106**, (2018).

60. Raimondi, F. *et al.* Fast-track virtual reality for cardiac imaging in congenital heart disease. *J. Card. Surg.* **36**, 2598–2602 (2021).

61. Ayerbe, V. M. C., Morales, M. L. V., Rojas, C. J. L. & Cortés, M. L. A. Visualization of 3D Models Through Virtual Reality in the Planning of Congenital Cardiothoracic Anomalies Correction: An Initial Experience. *World J. Pediatr. Congenit. Heart Surg.* **11**, 627–629 (2020).

62. Laas, E., El Beheiry, M., Masson, J.-B. & Malhaire, C. Partial breast resection for multifocal lower quadrant breast tumour using virtual reality. *BMJ Case Rep.* **14**, e241608 (2021).

63. Tomikawa, M. *et al.* Real-Time 3-Dimensional Virtual Reality Navigation System with Open MRI for Breast-Conserving Surgery. *J. Am. Coll. Surg.* **210**, 927–933 (2010).

64. Lang, H. & Huber, T. Virtual and Augmented Reality in Liver Surgery. *Ann. Surg.* **271**, e8 (2020).

65. Reitinger, B., Bornik, A., Beichel, R. & Schmalstieg, D. Liver Surgery Planning Using Virtual Reality. *IEEE Comput. Graph. Appl.* **26**, 36–47 (2006).

66. François, R. *et al.* Management of ischiopagus twin separation with a focus on W–S incision design. *J. Pediatr. Surg. Case Rep.* 101747 (2020)

doi:10.1016/j.epsc.2020.101747.

67. Verhey, J. T., Haglin, J. M., Verhey, E. M. & Hartigan, D. E. Virtual, augmented, and mixed reality applications in orthopedic surgery. *Int. J. Med. Robot.* **16**, (2020).

68. ImmersiveTouch®. *Medical VR Solutions*. (2021).

69. SurgicalTheater. *Experience 360® XR ANATOMY*. (2021).

frontiers in Bioinformatics

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fbinf.2021.777101

New Approach to Accelerated Image Annotation by Leveraging Virtual Reality and Cloud Computing

Corentin Guérinot[1,2,3†], Valentin Marcon[1†], Charlotte Godard[1,4†], Thomas Blanc[3,5], Hippolyte Verdier[1,6,7], Guillaume Planchon[1], Francesca Raimondi[1,8,9,10], Nathalie Boddaert[9,10], Mariana Alonso[2], Kurt Sailor[2], Pierre-Marie Lledo[2], Bassam Hajj[3,4‡], Mohamed El Beheiry[1‡] and Jean-Baptiste Masson[1*‡]*

[1]Decision and Bayesian Computation, USR 3756 (C3BI/DBC) & Neuroscience Department CNRS UMR 3751, Université de Paris, Institut Pasteur, Paris, France, [2]Perception and Memory Unit, CNRS UMR3571, Institut Pasteur, Paris, France, [3]Sorbonne Université, Collège Doctoral, Paris, France, [4]École Doctorale Physique en Île-de-France, PSL University, Paris, France, [5]Laboratoire Physico-Chimie, Institut Curie, PSL Research University, CNRS UMR168, Paris, France, [6]Histopathology and Bio-Imaging Group, Sanofi R&D, Vitry-Sur-Seine, France, [7]Université de Paris, UFR de Physique, Paris, France, [8]Unité Médicochirurgicale de Cardiologie Congénitale et Pédiatrique, Centre de Référence des Malformations Cardiaques Congénitales Complexes M3C, Hôpital Universitaire Necker-Enfants Malades, Université de Paris, Paris, France, [9]Pediatric Radiology Unit, Hôpital Universitaire Necker-Enfants Malades, Université de Paris, Paris, France, [10]UMR-1163 Institut Imagine, Hôpital Universitaire Necker-Enfants Malades, AP-HP, Paris, France

Three-dimensional imaging is at the core of medical imaging and is becoming a standard in biological research. As a result, there is an increasing need to visualize, analyze and interact with data in a natural three-dimensional context. By combining stereoscopy and motion tracking, commercial virtual reality (VR) headsets provide a solution to this critical visualization challenge by allowing users to view volumetric image stacks in a highly intuitive fashion. While optimizing the visualization and interaction process in VR remains an active topic, one of the most pressing issue is how to utilize VR for annotation and analysis of data. Annotating data is often a required step for training machine learning algorithms. For example, enhancing the ability to annotate complex three-dimensional data in biological research as newly acquired data may come in limited quantities. Similarly, medical data annotation is often time-consuming and requires expert knowledge to identify structures of interest correctly. Moreover, simultaneous data analysis and visualization in VR is computationally demanding. Here, we introduce a new procedure to visualize, interact, annotate and analyze data by combining VR with cloud computing. VR is leveraged to provide natural interactions with volumetric representations of experimental imaging data. In parallel, cloud computing performs costly computations to accelerate the data annotation with minimal input required from the user. We demonstrate multiple proof-of-concept applications of our approach on volumetric fluorescent microscopy images of mouse neurons and tumor or organ annotations in medical images.

Keywords: virtual reality, cloud computation, one-shot learning, inference, human-in-the-loop, MRI, CT-scan

OPEN ACCESS

Edited by:
Stephen Taylor,
University of Oxford, United Kingdom

Reviewed by:
Ondrej Strnad,
King Abdullah University of Science and Technology, Saudi Arabia
William C. Ray,
Nationwide Children's Hospital, United States

***Correspondence:**
Mohamed El Beheiry
mohamed.el-beheiry@pasteur.fr
Jean-Baptiste Masson
jbmasson@pasteur.fr

†These authors have contributed equally to this work and share first authorship

‡These authors have contributes equally to this work and share last authorship

Specialty section:
This article was submitted to Data Visualization, a section of the journal Frontiers in Bioinformatics

Received: *14 September 2021*
Accepted: *15 December 2021*
Published: *31 January 2022*

Citation:
Guérinot C, Marcon V, Godard C, Blanc T, Verdier H, Planchon G, Raimondi F, Boddaert N, Alonso M, Sailor K, Lledo P-M, Hajj B, El Beheiry M and Masson J-B (2022) New Approach to Accelerated Image Annotation by Leveraging Virtual Reality and Cloud Computing. Front. Bioinform. 1:777101. doi: 10.3389/fbinf.2021.777101

1 INTRODUCTION

Continuous technological advances in optical and electron microscopy have enhanced our ability to discern three-dimensional (3D) biological structures via slice-based tomography (Zheng et al., 2018; Driscoll et al., 2019; Gao et al., 2019; Hörl et al., 2019; Hoffman et al., 2020). Entire structures from organelles to whole organisms can be imaged at the nanometric resolution, allowing the exploration of complex interplay between 3D geometry and biological activity (Gao et al., 2019). Furthermore, large-scale recordings capturing entire organisms provide a new means for understanding biology at multiple spatial and temporal scales. Three-dimensional medical imaging has been accessible for many years (typically at the millimetric resolution), primarily acquired from computed tomography (CT) scans, magnetic resonance imaging (MRI), and, more recently, numerically processed ultrasound recordings. Medical image analysis is based on the specialized exploration of the slices along the principal axes of recording, i.e., the sagittal, coronal, and axial planes. These last 10 years have seen numerous machine learning-based approaches to assist and automate medical image analysis (Esteva et al., 2021).

Gaining an intuitive understanding from these complex raw data remains a challenge. Due to noise and statistical variability in the recordings, biological researchers often encounter difficulties in probing the geometry of organelles. It is also challenging in the medical imaging domain, where surgeons and clinicians lacking radiology training have difficulties in mentally transforming information in 2D image slices into a 3D representation of an organ, tumor or region of interest. In addition, natural modes of 3D visualization are missing, as most analyses rely on viewing 3D data on a computer monitor while simultaneously using a mouse to interact and extract information from the data.

Virtual reality (VR) technology has recently reemerged, in part due to low-cost consumer headsets and increasingly powerful graphics cards. The efficient integration of stereoscopy, immersion, and motion tracking in VR allows the user to visualize 3D structures in a physically realistic computer-generated environment. Interactions in this artificial environment rely on handheld VR controllers that allow physically-based actions to be performed on virtual objects.

Numerous initiatives have focused on taking advantage of this technology in the domains of education and scientific research (Dede et al., 2017; Balo et al., 2017; O'Connor et al., 2018; Johnston et al., 2018; Matthews, 2018; El Beheiry et al., 2019; Safaryan and Mehta, 2021). Recent studies have additionally highlighted the benefits of immersive viewing for handling 3D data, including efficiency and enhanced intuition relative to standard monitor-based visualization (Johnston et al., 2018; El Beheiry et al., 2019). It is worth pointing out that multiple companies have focused their efforts on developing state-of-the-art processes for high-quality image rendering. Examples of these active initiatives are found in Arivis AG and syGlass (see Table 1 in El Beheiry et al., 2020). Within the context of medical applications, initiatives have also focused on education (Djukic et al., 2013; Fertleman et al., 2018; Bouaoud et al., 2020; Shao et al., 2020; Venkatesan et al., 2021), surgery planning and diagnosis (Reitinger et al., 2006; Ong et al., 2018; Pfeiffer et al., 2018; Ayoub and Pulijala, 2019; Lee and Wong, 2019; Pinter et al., 2020; Wake et al., 2020; Boedecker et al., 2021; Chheang et al., 2021; Laas et al., 2021; Lau et al., 2021; Raimondi et al., 2021; Ruiz et al., 2021; Venkatesan et al., 2021).

Experimental three-dimensional image recordings (e.g., microscopy and medical) are typically acquired in limited quantities (Matthews, 2018). Additionally, these few acquisitions are subject to variability which make for difficult streamlining of data analysis. To address this reality we require, first, the appropriate means to visualize, interact with, and manipulate data and, second, an ability to rapidly perform quantitative assessments on these data.

The first challenge can be tackled via visualization with VR. By rendering image stacks into a VR environment, users can easily navigate and interact with their 3D data. In turn, VR enables the user to grasp an intuitive understanding of the dataset being visualized. However, multiple issues are associated with this task: 1) finding proper ways to represent diverse image stacks originating from different imaging modalities with varying signal to noise ratios, 2) providing versatile tools to explore and interact with the VR representation, and 3) finding procedures that can handle large data sets.

The second challenge is addressed by employing human-in-the-loop (Patel et al., 2019) data treatment procedures. The idea here is to couple user interactions with data analysis for extracting relevant information from the data. In the context of this work, this implies 1) defining procedures to select data within the VR environment, 2) performing the required computations for analysis without significantly impacting the VR rendering performance, and 3) allowing corrections to be performed in an iterative fashion.

In the following sections, we discuss related works involving VR software for image stack visualization. We introduce our approach, DIVA Cloud, which allows visualization and interaction in VR combined with cloud computing. Finally, we show how this approach can be effectively utilized in data annotation for microscopy and medical images.

2 QUICK INTRODUCTION TO RELATED WORKS

Affordable VR headsets, efficient graphics cards and easily accessible software development platforms (OpenXR, Unity, Unreal Engine etc.) have widened access to VR developments. These factors have promoted initiatives combining imaging techniques with VR in order to address topics in cell biology. As a result, VR applications are often forecasted to become essential components of the experimental research environment (Matthews, 2018; El Beheiry et al., 2019).

Image stack visualization in the VR environment is at the center of numerous initiatives. These include ConfocalVR (Stefani et al., 2018) and Scenery (Gunther et al., 2019), which can generate volumetric reconstructions of microscopy images. Neuroscience is a domain with a great need to visualize and manipulate data in 3D. Large images of entire nervous systems

can now be acquired with optical and electron microscopy (EM). Tracing complex neuronal structures is essential as there is a link between structure, connectivity and functions of neural circuits. Some initiatives already use VR to address this (Peng et al., 2010; Usher et al., 2017). Applications at the frontiers of microscopy and neurosurgery have also been demonstrated in literature (de Mauro et al., 2009; Wang et al., 2019; Wisotzky et al., 2019).

VR interactions from multiple users on the same data have been introduced in Naviscope (Shattuck, 2018). Additionally, some projects address topic-specific challenges in microscopy, such as colocalization (Theart et al., 2017).

Companies developing software for microscopy image analysis are now adding VR compatibility for visualization and treatment. Major advances are found with Arivis AG (Dekkers et al., 2019; Conrad et al., 2020) syGlass, which include an optimized data interaction interface. Other research applications focus on biomolecule structural information visualizations and interactions (Doutreligne et al., 2014; Balo et al., 2017; Goddard et al., 2018; Cassidy et al., 2020).

Not all microscopy image analysis software involves raw, full-stack image analyses but instead a deconvolved output (Betzig et al., 2006; Manzo and Garcia-Parajo, 2015; Qi et al., 2016). It is especially the case for single-molecule microscopy, where signals from individual biomolecules are captured and processed to deduce their nanometric positions and dynamic behavior. In these cases, the microscopy image stacks are reduced to point clouds. Two recent open-source software tools have been introduced to visualize and interact with single-molecule experiments: vLUME (Spark et al., 2020) and ours, Genuage (Blanc et al., 2020). Both software offers interfaces to interact with the point clouds and to perform various forms of data analysis (measuring, counting, cropping, etc.). Other initiatives on point clouds relate to data tagging for machine learning (Berge et al., 2016; Stets et al., 2018; Ramirez et al., 2019; Wirth et al., 2019; Liu et al., 2020). Mixed applications can be found involving astronomy, such as Gaia Sky (Sellés, 2013). General visualization and interaction software include PointCloud XR and developments centered on compression to ensure visualization in VR and Augmented Reality (AR) (Pavez et al., 2018).

In medicine, VR has found applications in surgery-specific topics, notably education. While VR may be useful for radiology (Uppot et al., 2019; Elsayed et al., 2020), radiologists are trained to perform 3D mental reconstructions of medical images, limiting their interest in immersive visualization modalities. Craniofacial trauma education (Bouaoud et al., 2020), neurosurgical training (Bernardo, 2017), spinal surgery (Pfandler et al., 2017), anatomy education (Uruthiralingam and Rea, 2020), orthopedic surgery (Bartlett et al., 2018; Walbron et al., 2019; Yoo et al., 2019; Lohre et al., 2020) and patient education (Dyer et al., 2018) have been demonstrated in this regard. Furthermore, clinical results hint towards uses of VR in surgical applications spanning heart diseases (Ayerbe et al., 2020; Sadeghi et al., 2020; Hattab et al., 2021; Raimondi et al., 2021), breast cancer (Tomikawa et al., 2010; Laas et al., 2021), liver surgery (Reitinger et al., 2006; Quero et al., 2019; Golse et al., 2020; Lang and Huber, 2020; Boedecker et al., 2021), pediatric surgery (Wang et al., 2012; Ruiz et al., 2021; Salvatore et al., 2021) and orthopaedic surgery (Bartlett et al., 2018; Yoo et al., 2019; Verhey et al., 2020). Multiple new companies are now investigating the potential of VR for surgical planning such as ImmersiveTouch®, PrecisionOS or SurgicalTheater.

3 VISUALIZING AND INTERACTING WITH IMAGE STACKS WITHOUT PRE-PROCESSING IN VR

We recently introduced DIVA software (El Beheiry et al., 2020), a user-friendly platform that generates volumetric reconstructions from raw 3D microscopy image stacks and enables efficient visualization, analysis and quantification. The software is available at https://diva.pasteur.fr.

DIVA was developed using the Unity game engine (UnityTechnologies), and is based on what we term a lean mapper software architecture. Furthermore, the software uses the Windows-based SteamVR standard, making it compatible with most PC VR headsets, such as the HTC Vive and Oculus Rift S. DIVA renders image stacks and hyperstacks instantaneously as 3D volumes through via GPU-based volume ray-casting (Engel et al., 2004, 2017). DIVA offers a dual interface allowing the user to interact both on a standard computer monitor (i.e., desktop mode) and in an immersive artificial environment (i.e., VR mode). However, spending a significant amount of time in VR can lead to discomfort among many users. Therefore, the desktop mode allows users to set optimal visualization parameters before switching to the VR mode, which is dedicated to visually interpreting, analyzing, and navigating the data.

In the desktop interface, the user can modify scaling and lighting, voxel color and opacity in real-time through a user-friendly transfer function interface. This transfer function allows a simple association of color and opacity to visualized voxels based on their raw intensity values, as shown in **Figure 1A**. Configuration of the transfer function and interaction with the volume (rotation, translation, and scaling) can be controlled with the mouse. Transfer functions in DIVA can be saved and loaded in Javascript Object Notation (JSON) format.

In DIVA's VR mode, the user visualizes image stacks rendered as live physical objects as a result of stereoscopy. Physical manipulation of the volume, such as grasping, rotation, or navigation, can be done with the VR controller, which acts as a 3D mouse for interaction (see **Figure 1B**). Hence, the data can be observed at any arbitrary angle to understand its structure in detail. To ease the navigation in complex and dense biological images, a clipping tool, presented in **Figure 1C**, can be activated. It dynamically removes a planar portion of the rendered volume allowing deep structures in the image to be revealed. Tagging, counting, and distance measurement tools are included for basic quantitative measurements (see **Figures 1D,E**). Users can extract all measurement results in a CSV file as well as through screen and movie captures.

As shown in **Figures 2A–C**, users can utilize DIVA for various microscopy modalities with up to four different channels associated with individual transfer functions. The user can

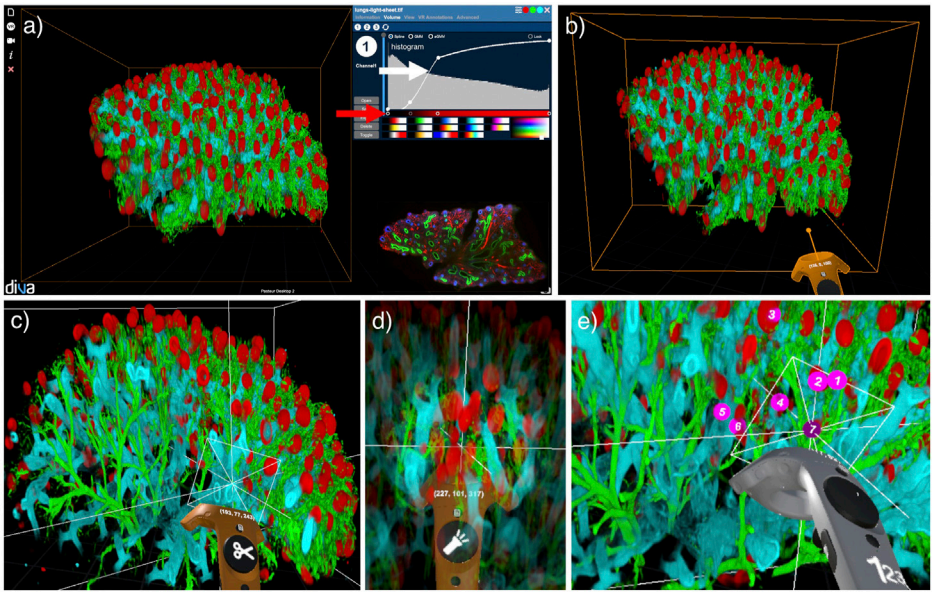


FIGURE 1 | The DIVA dual interface presented on an example of a light-sheet microscopy image of a human fetus lung with pulmonary alveoli (in red), trachea (in blue) and vascular system (in green) (Belle et al., 2017). **(A)** Desktop interface with raw data in the bottom right corner and transfer function interface in the top right corner with curves for voxel opacity (white arrow) and color (red arrow). **(B)** VR interface with VR controller in orange. **(C)** Clipping tool with the VR controller to navigate inside the volume. **(D)** Flashlight tool with the VR controller to highlight a spherical area of interest. **(E)** Counter tool with the VR controller to enumerate elements of interest.

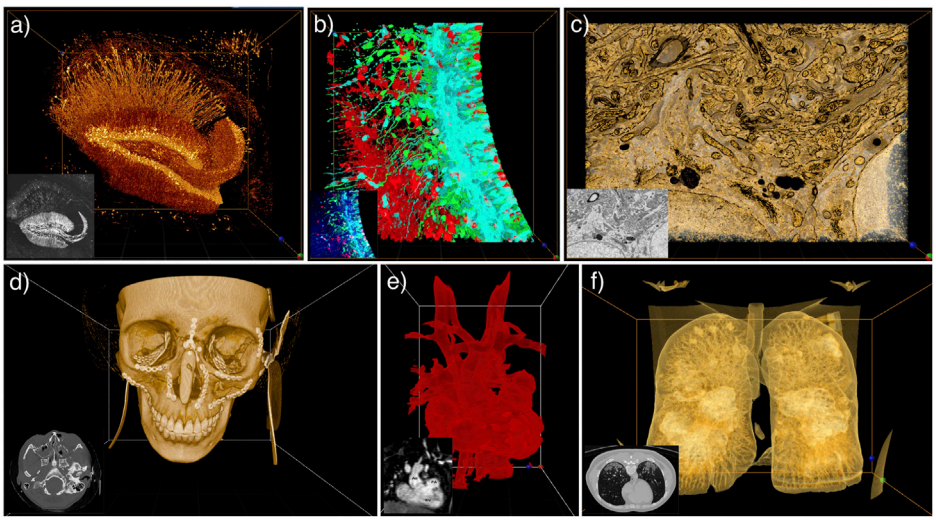


FIGURE 2 | DIVA application examples on the desktop interface with their corresponding raw image in the bottom left corner. **(A–C)** TIFF image stacks of **(A)** Mouse hippocampus imaged by two-photon serial endblock imaging (SEBI, Thy-1-GFP mouse) (Sun et al., 2013). **(B)** Mouse embryonic brain slices from spinning disk microscope (Brault et al., 2016). **(C)** Focused ion beam scanning EM of components of an adult mouse neuron: Golgi apparatus and mitochondria (Gao et al., 2019). **(D–F)** DICOM images of **(D)** Post CT-scan of craniofacial fractures (Bouaoud et al., 2020). **(E)** MRI of an adult heart with ventricular D-loop and septal defect (Raimondi et al., 2021). **(F)** CT-scan of lung with COVID-19 infection (Cohen et al., 2020).

reveal structures of interest by modulating the voxel transparency and colors in the transfer function. He can also discard undesired voxels without losing information, which can happen when using segmentation techniques. For example, users can easily remove the significant background noise of EM images. DIVA can also be used to compare raw and segmented data by merging the image stacks together as TIFF files with multiple channels.

With VR, the perception of 3D structures in complex data (e.g., EM images) is enhanced, and measurements are performed quicker than in standard 2D stack viewers. Examples of

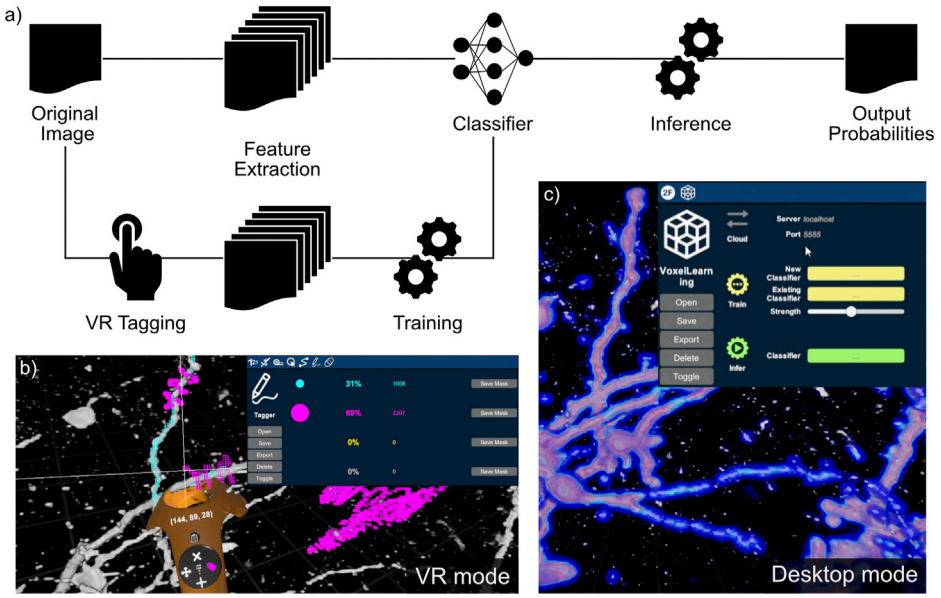


FIGURE 3 | Voxel Learning and its application on a confocal image stack of mouse olfactory bulb interneurons. **(A)** Schematic of the analysis pipeline. After having set visualization parameters, the user performs the VR tagging and selects the model to be used. Training and inference steps are performed on the cloud, as indicated by the pictogram. **(B)** Data tagging step with the VR controller in orange. The positive and negative tags are colored in cyan and magenta, respectively. **(C)** Voxel Learning interface in DIVA with the output probabilities overlaid on the original image (0 corresponds to blue; 1 to red).

advantages in 3D perception of VR are found for histological sample examination (Lobachev et al., 2021), surgery simulation or planning (Seymour, 2008; Guerriero et al., 2018; Thomsen et al., 2017; Chen et al., 2020), motion or gaze precision (Martirosov et al., 2021; Pastel et al., 2021), and 3D data labeling prior to machine learning training (Ramirez et al., 2020). Medical experts and undergraduate students have reported better visualization of 3D anatomical structures in VR using DIVA, when compared with typical 3D renderings, see Table 1 in Bouaoud et al. (2020) and Raimondi et al. (2021).

Medical images are most often stored in the Digital Imaging and Communications in Medicine (DICOM) format and analyzed through a 2D interface (i.e., a DICOM Viewer). In **Figures 2D–F**, we show examples of medical images visualized within DIVA. As DIVA is data agnostic, the user can experience both medical and microscopy images.

4 IMPLEMENTATION

In this work, we seek to use the DIVA VR visualization context with 3D image analysis. We focused on the acceleration of image annotation with the objective of exploit new data without prior information or a pre-trained model. We aimed at reducing the burden of data tagging, which can require a large amount of user interaction. Our procedure consists of rapid 3D tagging in VR, simple classifier training and inference on the entire dataset with iterative corrections performed within VR. The complete procedure, Voxel Learning, is described schematically in **Figure 3A** and in the included **Supplementary Video S1**.

4.1 Annotation in VR and Feature Extraction

In this updated version of the DIVA software, we implemented a VR tagging functionality which allows voxel annotation i.e., associating an identifier (ID) to individual voxels. Tagged voxels appear in different colors, in a transparent wireframe mesh around the tagged voxel, allowing simultaneous visualization of the voxel and its tag. In the application shown in **Figure 3B**, the two colors are cyan and magenta as they highly visible in most situations. Tags can be updated or erased if necessary. The clipping plane tool (CPT) (El Beheiry et al., 2020) is also available to ensure more precision in ambiguous situations (see **Figure 1C**) and fluid tagging within the volume. Most importantly, the CPT allows annotating data at the frontier between different domains with geometries that do not align along the natural axis of data acquisition. Inside the VR environment, the properties of the interface and the transfer function are instrumental in accelerating the annotation process. We demonstrate the tagging procedure in **Supplementary Video S2** on a medical example.

We additionally integrated the ability to calculate image features with the DIVA software. An efficient feature evaluation was implemented for each voxel, using a small subset of features (Arganda-Carreras et al., 2017). It includes a wide variety of spatial filters (Gaussian, median, mean, etc.) with different kernel sizes for gathering simple multi-scale features in the vicinity of the voxels. Features are then associated with a unique voxel ID, and the list of annotated voxels is stored. In the case of iterative tagging, the iteration number is also stored and associated with the voxel IDs.

4.2 Training and Inference

Our application here consists of accelerating data annotation using a simple one-shot learning procedure based on a few VR controller "strokes" performed by the user on the image stack. We follow the same principles as those used in ilastik (Berg et al., 2019), Weka (Arganda-Carreras et al., 2017) or behavior detection in larva (Masson et al., 2020), by tagging limited sets of data and stacking simple learners in order to train a collectively stronger classifier. Specifically, features are associated with tagged voxels, and learning is performed using robust classifiers in limited amounts of data. Furthermore, data tagging iterations allow the correction of anomalies in the learning to process the data.

We updated the DIVA desktop user interface to allow users to create quickly, load, and export classifiers. Basic classification approaches were used since they are known to provide robust classification on small datasets, such as Random Forest Classification (RFC), Multi-Layer Perceptron (MLP), Gradient Boosting (XGB), Support Vector Machine (SVM), and Naive Bayes (NBC) as implemented in the Scikit-learn (Pedregosa et al., 2011) Python package. In addition, hyperparameters were tuned to adapt to the problem being investigated and set to ensure rapid learning. Note that in this application, the usual problem of overfitting (Mostafa, 2012) is less present, as the goal is to annotate the data being explored and not to find a general learning scheme. Once the user has finished the annotation step, features associated to voxels are evaluated locally. They are then transferred in JSON format to the cloud, where a model is trained to classify all voxels in the 3D data stack. The models and their associated parameters are saved locally in a Pickle (Python) format. They can afterwards be loaded to perform inference on the entire dataset or, if found robust, provide initial annotations on new, previously unseen, data. The resulting inference is then broadcasted back from the cloud for local rendering in DIVA.

VR provides a significant advantage in the data annotation task by properly overlaying the classification result on the raw data in a volumetric representation. The representation of both the raw data and the annotated data provided by the classifiers helps correct errors and ensures proper annotation. We integrated a channel-based representation to DIVA in which raw data and classifier-generated data are associated with different channels allowing separate and fused views of both raw and classified data. A transfer function interface is associated for each channel. In most of our applications, the raw data was fused with the voxel probability (or log-probability) of belonging to a specific class. An example is shown in **Figure 3C** and in **Supplementary Video S3**.

Overall classifier robustness can be improved by using an ensemble of weak classifiers, whose resulting probabilities are added to the list of features before final model training. In addition, classifiers may also be iteratively updated with additional tagging rounds to correct for sub-optimal performance and false detection. Stacked learners are efficient in performing intuitive segmentation (Sommer et al., 2011). We denote here the gradient boost classifier with four weak learners as the strong learner.

4.3 DIVA Cloud

DIVA Cloud allows users to interact with data in VR and perform the analysis via Python scripts (see **Section 6**) whose calculations are performed on a remote web server. This development is motivated by the computational costs required for detailed VR volumetric rendering, leaving limited calculation bandwidth. Accordingly, to ensure fluid interactions and precise annotation dynamics, computationally costly operations are performed on the remote web server (i.e. cloud). We used Django, a Python Framework, as an application programming interface (API) provider for this project. Django features a system of data models and serializer links to a PostgreSQL database, enabling the management of jobs and file objects to track the life cycle of the learning jobs and various input and output files. This system is instrumental, as it permits performing a limited set of queries to interact with the cloud component.

The REST web service is used to provide an API with specific endpoints accessible from DIVA using HTTP methods. Celery, a task queue intermediate between the web server and the Python scripts, enables jobs to run asynchronously in a multi-threaded fashion. At the same time, the broker, Redis, allows communication with Celery.

We implemented DIVA Cloud within a Docker container to ensure portability on most platforms since it does not depend on the user installation. The layers implemented in the container consist of 1) the Django webserver, 2) the Celery task queue that integrates learning scripts (with their associated packages), 3) the Redis broker, and 4) the PostgreSQL database. Note that if the user runs DIVA Cloud on a powerful computation station, the entire pipeline can be executed locally on the same computer already running DIVA. A typical interaction workflow for the DIVA Cloud application is shown in **Figure 4A** in the context of data tagging. A visual representation of the relation between DIVA and DIVA Cloud is shown in **Figure 4B**.

5 RESULTS

Our analysis was performed on a Windows 10 based ×64 system with an Intel i7-7700 CPU clocked at 3.60 GhZ, with 32 GB of RAM and an NVIDIA GeForce RTX 2080 Ti graphics card. An HTC Vive headset with its controller were used through SteamVR to perform the tagging procedure in VR. This particular VR headset has a total screen resolution of 2,160 × 1,200 pixels (1,080 × 1,200 for each eye). Analysis scripts were coded in Python 3.7. The DIVA Cloud configuration was also tested on a NVIDIA DGX-1 workstation as the remote server performing the computationally challenging tasks. The frame rate of DIVA Cloud is highly dependent on the hardware configuration, image size, interface (Desktop interface in 2D or VR interface), and the user's movements in VR. Detailed information on the frame rate regarding this study can be found in **Supplementary Table S3**.

5.1 Metrics

We showed a proof of concept of this approach on various example image stacks, including a CT-scan, MRI sequence and various microscopy images applied to neuronal specimens.

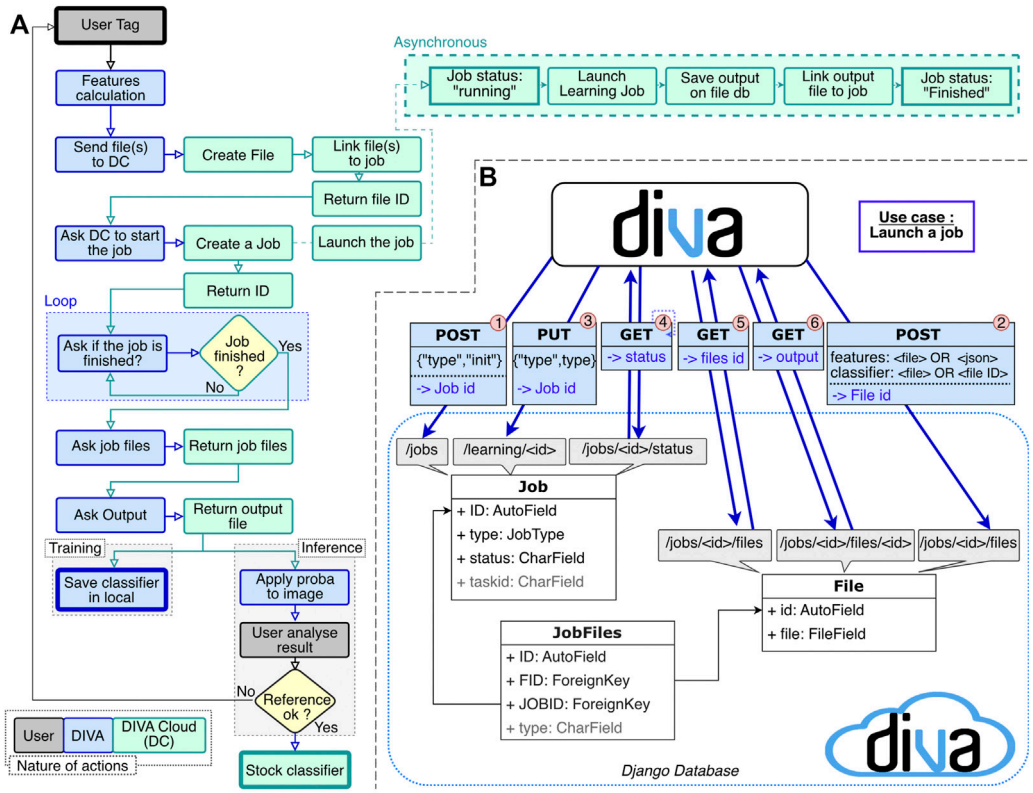


FIGURE 4 | (A) DIVA Cloud interaction workflow through a data tagging experiment to output a classifier that is visualized in DIVA. **(B)** Interaction between DIVA and DIVA Cloud in 6 steps: 1) POST request to the/jobs endpoint. It initializes a job entry in Django. Get in return the job ID 2) POST request to the/jobs/*jobid*/file endpoint with the inputs files. It creates a file entry and returns the file ID 3) PUT request to the/learning/*jobid* endpoint with the type of learning. It launches the job on input data 4) GET request to the/jobs/*jobid*/status endpoint to know the status of the job. If the status is "running", the status of the job is requested (Step 4 again). If the status is "done", continue. If the status is "error", it is managed. 5) GET request to the/jobs/*jobid*/files/endpoint to get the output list 6) GET request to the/jobs/*jobid*/files/*fileid* to download the output.

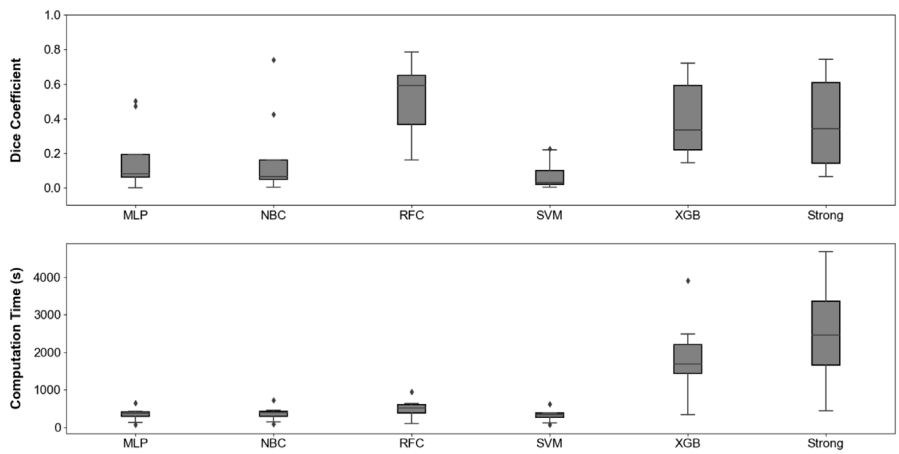


FIGURE 5 | Distribution of Dice coefficient **(A)** and computation time **(B)** when applying our annotation procedure to eight different medical examples images. Corresponding raw data is available in **Supplementary Table S1**.

We kept track of different time measurements: the tagging step in VR never exceeded 2 min. Tagging and subsequent model training are performed on a small portion of the data. Inference duration scaled with the size of the 3D image and depended on the computational bandwidth of the cloud infrastructure.

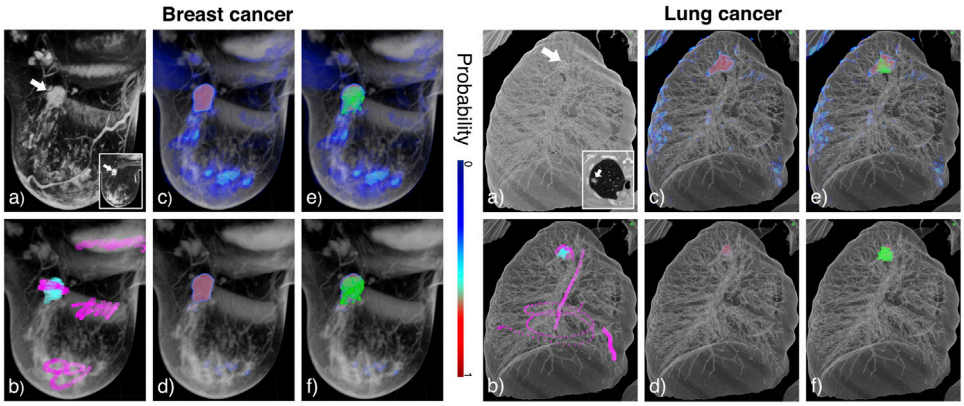


FIGURE 6 | Annotation in DIVA on the breast MRI (left panel) and the lung CT-scan (right panel) and tumor (white arrow). **(A)** Raw data visualized in 3D on DIVA and as an image stack in the bottom right corner. **(B)** Overlay of the raw image in gray and tags with positive and negative tags in cyan and magenta, respectively. Tagging is performed in VR to quickly annotate which voxels belong to the structure of interest and which do not. **(C,D)** Overlay of the raw image in gray and output probabilities, respectively for the RFC and the strong learner. **(E,F)** Overlay of the raw image in gray, output probabilities, and ground truth segmentation in green for RFC and strong learner, respectively. Colorscale for probabilities is indicated between the two panels.

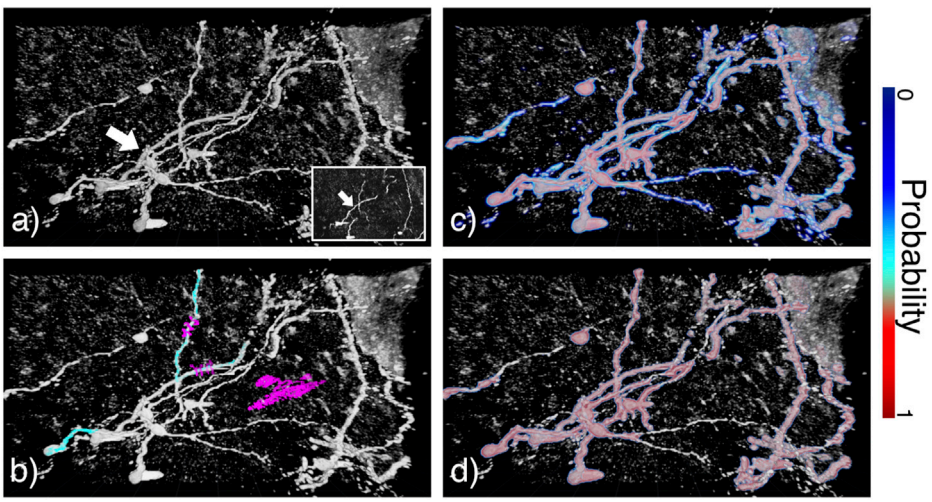


FIGURE 7 | Annotation in DIVA on confocal microscopy images of mouse olfactory bulb interneurons (white arrow). **(A)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **(B)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **(C,D)** Overlay of raw data in gray and output probabilities respectively for RFC and strong learner. Colorscale for probabilities is indicated on the right of the image.

For medical images, we used raw and annotated images in order to compare the one-shot annotation to the expert full tagging. We computed Dice coefficient and Residual Mean Square Errors (RMSE) between our inferred probabilities and the given segmentation. The goal here was to evaluate how fast annotation in VR and quick simple learning can reduce the tagging of new data to a few complementary VR strokes. Performance depends on the nature of the features extracted from the dataset. As features were designed to cover a large variety of patterns and scales, our method can see use in many additional applications and data types. Furthermore, the number and nature of the features can be extended to capture specific properties of image stacks. All measurements are available in **Supplementary Tables S1 and S2**.

Figure 5 compares the Dice coefficient obtained with different models, along with the corresponding computing time when applied to medical examples images. The RFC and XGB, when stacked with 4 weak learners, reached highest performance with the Dice coefficient. RFC is associated to a shorter computation time, making it an appropriate candidate for efficient analysis.

5.2 Output Probabilities

In order to compare the results of RFC and the strong learner, we show in **Figure 6** their application to MRI images showing a

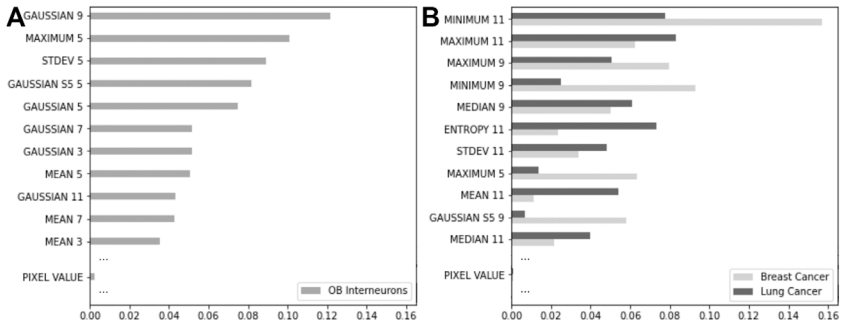


FIGURE 8 | Feature importance for the RFC trained on a confocal image of the mouse olfactory bulb interneurons **(A)**, and the MRI of a breast cancer and CT-scan of a lung cancer **(B)**. The top eleven features ranked by impurity-based importance are represented, as well as the PIXEL VALUE feature.

patient with breast cancer and a CT-scan of a patient with lung cancer (see **Supplementary Figures S1, S2** for more examples). **Figure 6B** exhibits the tagging step. We decided to tag the lung tumor somewhat less completely than the breast tumor in order to assess the impact of tagging exhaustiveness. Regardless, our pipeline demonstrates qualitative results, identifying the structures of interest with precision using both models (see **Figures 6C,D**). We note the presence of low probability inferred voxels ranging from dark to light blue for RFC, whereas the strong learner seems to classify in a more binary fashion. In this respect, RFC proves to be more prone to false positive detection. The strong learner was more stringent and, as a result, may induce more false negative errors. We observe this tendency in **Figures 6E,F**, as RFC predictions exceed ground truth segmentation in volume while the strong learner predictions appeared smaller.

We also tested our accelerated data annotation procedure on microscopy images. Mouse olfactory bulb interneurons were imaged *via* confocal microscopy, results are shown in **Figure 7**. These data were considerably noisier than the CT-scan and MRI sequences that were previously utilized. The objective of this analysis was to reconstruct neuronal dendritic arbors. This was achieved by tagging two neuronal branches (see **Figure 7B**) and then applying our pipeline using the RFC and strong learner (see **Figures 7C,D** respectively). Almost every neurite and soma were classified using both learners. The RFC yielded high probabilities for inner structures and lower ones for outer structures, allowing isolating tubular structures through proper adaptation of the transfer function.

The procedure was assessed on a various range of examples: annotation of pancreas (**Supplementary Figures S3, S4**) and hepatic vessels (**Supplementary Figures S5, S6**) in CT-scans, annotation of mouse microglia in two-photon fluorescence microscopy (**Supplementary Figure S7**) and mouse hippocampal neurons in SEBI microscopy (**Supplementary Figures S8, S9**). With a total size of more than 400 million pixels, this last example confirms that our pipeline is compatible with large datasets.

5.3 Feature Importance

The RFC enables a ranking of the features used during model training through a metric called impurity-based importance. The importance of a feature is calculated as the reduction in its Gini criterion. This reduction in impurity is afterward express in percentages. We show in **Figure 8** the eleven most important features for the three examples previously presented, as well as the importance of *PIXEL VALUE* corresponding simply to the intensity of a given voxel (see **Supplementary Figures S10, S11** for the complete features importance for the 56 features used). Interestingly, this feature, generally used in thresholding for crude denoising or segmentation, seems to have almost no importance in the final prediction.

Comparing the RFC impurity-based importance from both neuroscience microscopy and medical image types (CT-scans and MRI), we note a large variety of features that drive the final classification prediction. Indeed, for medical images, it seems that large spatial filters with 3D kernel sizes of 9–11 voxels contribute mainly to the importance, while the typical kernel size seems closer to 5 or 7 for microscopy. It can be explained by the different characteristics of the respective structures of interest (bulkier for the tumors and more tubular for the neuronal branches). While not surprising, an extended version of this procedure with larger sets of features can guide machine learning procedures that attempt to reduce the number of features used for learning.

6 DISCUSSION

Virtual and augmented reality will likely play an increasingly important role in research and medical applications. More specifically, leveraging VR as part of analysis pipelines will be essential in defining algorithms and processes.

This paper focused on performing interaction and analysis in VR. We developed libraries and toolboxes allowing data annotation and analysis where computationally intensive operations are done on servers or the cloud and data fusion within the VR environment. Our approach combines several updates of the DIVA platform and a generalist interface allowing cloud computing from the VR environment. We showed proof-of-concept results on an accelerated image annotation task in which new data requires human-in-the-loop intervention to provide initial results. This provides a direct process in which limited data annotation is sufficient to

train a simple statistical learning approach to classify voxels in entire image stacks.

In this work, we demonstrate the usefulness of VR in ameliorating the efficiency of data annotation tasks. The ability to grasp the volume of the data in a stereoscopic 3D viewing context and tag a small portion of it to perform an entire segmentation task is unique and promising. In contrast, the same task applied to 2D slices would require significant exploration and proves to be a tedious process.

We take advantage of the possibility of adjusting the transfer function to visualize the region of interests and their vicinity correctly. Rapidly tagging voxels not belonging to the structure of interest, by a large stroke using the VR controller (e.g., the pink streaks in **Figure 6**), will be instrumental in accelerating data annotation by easing learning. When compared to ground truths, the inspection of learned results is often a time-consuming task, especially when dealing with 3D data. Through its capacity to fuse different volumes in the representation, DIVA offers a suitable and time-saving environment to perform such a comparison. In turn, it allows exploring the reliability of the learning procedure and assess the quality of the ground truth itself when dealing with ambiguous data. It provides the possibility of re-annotation where anomalies are detected to feed the re-training procedure.

In this paper, we focused on the handcrafted procedure and simple ensemble learning approaches. While some of the data shown were noisy (e.g., **Figure 7**) or subject to artifacts (e.g., **Figures 6A–F**), they were relatively unambiguous. Local handcrafted features were sufficient to allow efficient semi-automated annotation. Furthermore, we limited ourselves to cases where data were completely new, and the learner would be mainly used on the explored dataset. However, complex data or the design of reusable learners may require learned features. The current pipeline can perform more complex learning. We showed (see https://github.com/DecBayComp/VoxelLearning) an example where VR annotation was directly used to capture a volume of voxels around the annotated ones and where learning was directly performed on these volumes with learned features. Finally, large-scale deep convolution-based learning can be performed by directly transmitting the full data to the cloud and using DIVA Cloud to link annotation to learning by simply exchanging voxel identifiers.

Our future initiatives will center on two topics. First, we will 1) extend DIVA and DIVA Cloud to run on AR headsets and tablets. AR involves overlaying the visual representation of the data onto the world while not immersing the user in an artificial environment. However, AR often involves representation of lesser quality than VR when used in glasses or headsets and have limited computational resources, especially when used in phones or tablets. The initiative's core is to reduce the computational burden of full volumetric rendering to allow visualization and interaction in an AR environment. Second, we will 2) extend DIVA Cloud to allow rendering computation on the cloud and stream to the VR headset. This cloud extension will pave the way to large dataset rendering and more computationally involved rendering approaches, such as path tracing.

DATA AVAILABILITY STATEMENT

Publicly available datasets and original contributions were analyzed in this study. This data as well as the entire pipeline can be found here: https://github.com/DecBayComp/VoxelLearning.

AUTHOR CONTRIBUTIONS

CGu, VM, CGo, TB, GP, MB, and J-BM contributed to the code of DIVA Cloud. VM was the sole developer of the Django interface of DIVA Cloud. MA, KS, and P-ML provided data related to neuroscience, performed data annotation and contributed to the research design. FR and NB provided data and medical expertise regarding visualization and data annotation. They contributed to paper revisions. BH, J-BM, and MB designed and supervised the research. CGu, VM, CGo, BH, MB, and J-BM wrote the paper.

FUNDING

We acknowledge funding from the Institut Pasteur, the Institut Curie, the Paris Science lettre (PSL) University, the sponsorships of CRPCEN, Gilead Science and foundation EDF, the INCEPTION project (PIA/ANR-16-CONV-0005, OG), the LabEx "Revive" (ANR-10-LABX-73), the programme d'investissement d'avenir supported by L'Agence Nationale de la Recherche ANR-19-P3IA-0001 Institut 3IA Prairie and the support of the AVIRON grant from the Région Ile-de-France (DIM-ELICIT). We acknowledge funding from the Fondation pour la recherche médicale (FRM; DEI20151234398), the Agence National de la recherche (ANR-19-CE42-0003-01), the LabEx CelTisPhyBio (ANR-10-LBX-0038, ANR-10-IDEX-0001-02) and the Institut Curie. We recognize the support of France-BioImaging infrastructure grant ANR-10-INBS-04 (Investments for the future). We acknowledge the financial support of the Agence pour la Recherche sur le Cancer (ARC Foundation), DIM ELICIT and from ITMO Cancer of Aviesan on funds administered by Inserm (grant No. 20CP092-00).

ACKNOWLEDGMENTS

We thank Alexandre Baffet (Institut Curie), Thomas Rubin and Jean-Ren Huynh (Collège de France) for the microscopic image datasets provided to generate the figures. We also thank Jebrane Bouaoud (AP-HP) for the medical images provided to generate the figures. We acknowledge Stephane Fournier, the naviscope Inria project lab teams & Thomas Isenberg for fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbinf.2021.777101/full#supplementary-material

REFERENCES

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., et al. (2017). Trainable Weka Segmentation: a Machine Learning Tool for Microscopy Pixel Classification. *Bioinformatics* 33, 2424–2426. doi:10.1093/bioinformatics/btx180

Arivis, A. G. (2021). *Image Visualization and Analysis*.

Ayerbe, V. M. C., Morales, M. L. V., Rojas, C. J. L., and Cortés, M. L. A. (2020). Visualization of 3d Models through Virtual Reality in the Planning of Congenital Cardiothoracic Anomalies Correction: An Initial Experience. *World J. Pediatr. Congenit. Heart Surg.* 11, 627–629. doi:10.1177/2150135120923618

Ayoub, A., and Pulijala, Y. (2019). The Application of Virtual Reality and Augmented Reality in Oral & Maxillofacial Surgery. *BMC Oral Health* 19, 238. doi:10.1186/s12903-019-0937-8

Balo, A. R., Wang, M., and Ernst, O. P. (2017). Accessible Virtual Reality of Biomolecular Structural Models Using the Autodesk Molecule Viewer. *Nat. Methods* 14, 1122–1123. doi:10.1038/nmeth.4506

Bartlett, J. D., Lawrence, J. E., Stewart, M. E., Nakano, N., and Khanduja, V. (2018). Does Virtual Reality Simulation Have a Role in Training Trauma and Orthopaedic Surgeons? *Bone Jt. J* 100-B, 559–565. doi:10.1302/0301-620x.100b5.bjj-2017-1439

Belle, M., Godefroy, D., Couly, G., Malone, S. A., Collier, F., Giacobini, P., et al. (2017). Tridimensional Visualization and Analysis of Early Human Development. *Cell* 169, 161–e12. doi:10.1016/j.cell.2017.03.008

Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., et al. (2019). Ilastik: Interactive Machine Learning for (Bio)image Analysis. *Nat. Methods* 16, 1226–1232. doi:10.1038/s41592-019-0582-9

Berge, L.-P., Aouf, N., Duval, T., and Coppin, G. (2016). "Generation and VR Visualization of 3d point Clouds for Drone Target Validation Assisted by an Operator," in Proceeding of the 2016 8th Computer Science and Electronic Engineering (CEEC), Colchester, UK, 28-30 Sept. 2016 (IEEE), 66–70. doi:10.1109/ceec.2016.7835890

Bernardo, A. (2017). Virtual Reality and Simulation in Neurosurgical Training. *World Neurosurg.* 106, 1015–1029. doi:10.1016/j.wneu.2017.06.140

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., et al. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* 313, 1642–1645. doi:10.1126/science.1127344

Blanc, T., El Beheiry, M., Caporal, C., Masson, J. B., and Hajj, B. (2020). Genuage: Visualize and Analyze Multidimensional Single-Molecule point Cloud Data in Virtual Reality. *Nat. Methods* 17, 1100–1102. doi:10.1038/s41592-020-0946-1

Boedecker, C., Huettl, F., Saalfeld, P., Paschold, M., Kneist, W., Baumgart, J., et al. (2021). Using Virtual 3d-Models in Surgical Planning: Workflow of an Immersive Virtual Reality Application in Liver Surgery. *Langenbecks Arch. Surg.* 406, 911–915. doi:10.1007/s00423-021-02127-7

Bouaoud, J., El Beheiry, M., Jablon, E., Schouman, T., Bertolus, C., Picard, A., et al. (2021). DIVA, a 3d Virtual Reality Platform, Improves Undergraduate Craniofacial Trauma Education. *J. Stomatology, Oral Maxillofacial Surg.* 122, 367–371. doi:10.1016/j.jormas.2020.09.009

Brault, J. B., Khou, C., Basset, J., Coquand, L., Fraisier, V., Frenkiel, M. P., et al. (2016). Comparative Analysis between Flaviviruses Reveals Specific Neural Stem Cell Tropism for Zika Virus in the Mouse Developing Neocortex. *EBioMedicine* 10, 71–76. doi:10.1016/0003-4916(63)90068-X10.1016/j.ebiom.2016.07.018

Cassidy, K. C., Šefčík, J., Raghav, Y., Chang, A., and Durrant, J. D. (2020). ProteinVR: Web-Based Molecular Visualization in Virtual Reality. *Plos Comput. Biol.* 16, e1007747. doi:10.1371/journal.pcbi.1007747

Celery (2021). *Distributed Task Queue — Celery 5.1.2 Documentation*.

Chen, E., Luu, W., Chen, R., Rafik, A., Ryu, Y., Zangerl, B., et al. (2020). Virtual Reality Improves Clinical Assessment of the Optic Nerve. *Front. Virtual Real.* 1, 4. doi:10.3389/frvir.2020.00004

Chheang, V., Saalfeld, P., Joeres, F., Boedecker, C., Huber, T., Huettl, F., et al. (2021). A Collaborative Virtual Reality Environment for Liver Surgery Planning. *Comput. Graphics* 99, 234–246. doi:10.1016/j.cag.2021.07.009

Cohen, J. P., Morrison, P., and Dao, L. (2020). *COVID-19 Image Data Collection*.

Conrad, R., Ruth, T., Löffler, F., Hadlak, S., Konrad, S., Götze, C., et al. (2020). Efficient Skeleton Editing in a VR Environment Facilitates Accurate Modeling of Highly Branched Mitochondria. *Microsc. Microanal* 26, 1158–1161. doi:10.1017/s1431927620017158

de Mauro, A., Raczkowsky, J., Halatsch, M. E., and Worn, H. (2009). "Virtual Reality Training Embedded in Neurosurgical Microscope," in In Proceeding of the 2009 IEEE Virtual Reality Conference, Lafayette, LA, USA, 14-18 March 2009 (IEEE), 233–234. doi:10.1109/vr.2009.4811031

Dede, C. J., Jacobson, J., and Richards, J. (2017). *Introduction: Virtual, Augmented, and Mixed Realities in Education*. Singapore: Springer.

Dekkers, J. F., Alieva, M., Wellens, L. M., Ariese, H. C. R., Jamieson, P. R., Vonk, A. M., et al. (2019). High-resolution 3d Imaging of Fixed and Cleared Organoids. *Nat. Protoc.* 14, 1756–1771. doi:10.1038/s41596-019-0160-8

Django (2016). *Django the Web Framework for Perfectionists with Deadlines*.

Djukic, T., Mandic, V., and Filipovic, N. (2013). Virtual Reality Aided Visualization of Fluid Flow Simulations with Application in Medical Education and Diagnostics. *Comput. Biol. Med.* 43, 2046–2052. doi:10.1016/j.compbiomed.2013.10.004

Doutreligne, S., Cragnolini, T., Pasquali, S., Derreumaux, P., and Baaden, M. (2014). "Unitymol: Interactive Scientific Visualization for Integrative Biology" in Proceeding of the 2014 IEEE 4th Symposium on Large Data Analysis and Visualization (LDAV), Paris, France, 9-10 Nov. 2014 (IEEE), 109–110. doi:10.1109/LDAV.2014.7013213

Driscoll, M. K., Welf, E. S., Jamieson, A. R., Dean, K. M., Isogai, T., Fiolka, R., et al. (2019). Robust and Automated Detection of Subcellular Morphological Motifs in 3d Microscopy Images. *Nat. Methods* 16, 1037–1044. doi:10.1038/s41592-019-0539-z

Dyer, E., Swartzlander, B. J., and Gugliucci, M. R. (2018). Using Virtual Reality in Medical Education to Teach Empathy. *J. Med. Libr. Assoc.* 106, 498–500. doi:10.5195/jmla.2018.518

El Beheiry, M., Doutreligne, S., Caporal, C., Ostertag, C., Dahan, M., and Masson, J. B. (2019). Virtual Reality: Beyond Visualization. *J. Mol. Biol.* 431, 1315–1321. doi:10.1016/j.jmb.2019.01.033

El Beheiry, M., Godard, C., Caporal, C., Marcon, V., Ostertag, C., Sliti, O., et al. (2020). DIVA: Natural Navigation inside 3d Images Using Virtual Reality. *J. Mol. Biol.* 432, 4745–4749. doi:10.1016/j.jmb.2020.05.026

Elsayed, M., Kadom, N., Ghobadi, C., Strauss, B., Al Dandan, O., Aggarwal, A., et al. (2020). Virtual and Augmented Reality: Potential Applications in Radiology. *Acta Radiol.* 61, 1258–1265. doi:10.1177/0284185119897362

Engel, K., Hadwiger, M., Kniss, J. M., Lefohn, A. E., Salama, C. R., and Weiskopf, D. (2004). "Real-time Volume Graphics," in *Proceedings of the Conference on SIGGRAPH 2004 Course Notes - GRAPH '04* (Los Angeles, CA: Association for Computing Machinery Press), 29–es. doi:10.1145/1103900.1103929

Engel, K., Hadwinger, M., Kniss, J. M., Rezk-Salama, C., and Weiskopf, D. (2017). *Real-Time Volume Graphics*. Wellesley, Massachusetts, USA: A K Peters, Ltd.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., et al. (2021). Deep Learning-Enabled Medical Computer Vision. *Npj Digit. Med.* 4. doi:10.1038/s41746-020-00376-2

Fertleman, C., Aubugeau-Williams, P., Sher, C., Lim, A. N., Lumley, S., Delacroix, S., et al. (2018). A Discussion of Virtual Reality as a New Tool for Training Healthcare Professionals. *Front. Public Health* 6, 44. doi:10.3389/fpubh.2018.00044

Gao, R., Asano, S. M., Upadhyayula, S., Pisarev, I., Milkie, D. E., Liu, T. L., et al. (2019). Cortical Column and Whole-Brain Imaging with Molecular Contrast and Nanoscale Resolution. *Science* 363, eaau8302. doi:10.1126/science.aau8302

Goddard, T. D., Brilliant, A. A., Skillman, T. L., Vergenz, S., Tyrwhitt-Drake, J., Meng, E. C., et al. (2018). Molecular Visualization on the Holodeck. *J. Mol. Biol.* 430, 3982–3996. doi:10.1016/j.jmb.2018.06.040

Golse, N., Petit, A., Lewin, M., Vibert, E., and Cotin, S. (2020). Augmented Reality during Open Liver Surgery Using a Markerless Non-rigid Registration System. *J. Gastrointest. Surg.* 25, 662–671. doi:10.1007/s11605-020-04519-4

Guerriero, L., Quero, G., Diana, M., Soler, L., Agnus, V., Marescaux, J., et al. (2018). Virtual Reality Exploration and Planning for Precision Colorectal Surgery. *Dis. Colon Rectum* 61, 719–723. doi:10.1097/DCR.0000000000001077

Gunther, U., Pietzsch, T., Gupta, A., Harrington, K. I. S., Tomancak, P., Gumhold, S., et al. (2019). "Scenery: Flexible Virtual Reality Visualization on the Java VM," in Proceeding of the 2019 IEEE Visualization Conference (VIS), Vancouver, BC, Canada, 20-25 Oct. 2019 (IEEE), 1–5. doi:10.1109/visual.2019.8933605

Hattab, G., Hatzipanayioti, A., Klimova, A., Pfeiffer, M., Klausing, P., Breucha, M., et al. (2021). Investigating the Utility of VR for Spatial Understanding in Surgical Planning: Evaluation of Head-Mounted to Desktop Display. *Sci. Rep.* 11, 13440. doi:10.1038/s41598-021-92536-x

Hoffman, D. P., Shtengel, G., Xu, C. S., Campbell, K. R., Freeman, M., Wang, L., et al. (2020). Correlative Three-Dimensional Super-resolution and Block-Face Electron Microscopy of Whole Vitreously Frozen Cells. *Science* 367, eaaz5357. doi:10.1126/science.aaz5357

Hörl, D., Rojas Rusak, F., Preusser, F., Tillberg, P., Randel, N., Chhetri, R. K., et al. (2019). BigStitcher: Reconstructing High-Resolution Image Datasets of Cleared and Expanded Samples. *Nat. Methods* 16, 870–874. doi:10.1038/s41592-019-0501-0

ImmersiveTouch (2021). *Medical VR Solutions*.

Johnston, A. P. R., Rae, J., Ariotti, N., Bailey, B., Lilja, A., Webb, R., et al. (2018). Journey to the centre of the Cell: Virtual Reality Immersion into Scientific Data. *Traffic* 19, 105–110. doi:10.1111/tra.12538

Laas, E., El Beheiry, M., Masson, J. B., and Malhaire, C. (2021). Partial Breast Resection for Multifocal Lower Quadrant Breast Tumour Using Virtual Reality. *BMJ Case Rep.* 14, e241608. doi:10.1136/bcr-2021-241608

Lang, H., and Huber, T. (2020). Virtual and Augmented Reality in Liver Surgery. *Ann. Surg.* 271, e8. doi:10.1097/sla.0000000000003601

Lau, I., Gupta, A., and Sun, Z. (2021). Clinical Value of Virtual Reality versus 3d Printing in Congenital Heart Disease. *Biomolecules* 11, 884. doi:10.3390/biom11060884

Lee, C., and Wong, G. K. C. (2019). Virtual Reality and Augmented Reality in the Management of Intracranial Tumors: A Review. *J. Clin. Neurosci.* 62, 14–20. doi:10.1016/j.jocn.2018.12.036

Liu, W., Lai, B., Wang, C., Bian, X., Yang, W., Xia, Y., et al. (2020). "Learning to Match 2d Images and 3d LiDAR point Clouds for Outdoor Augmented Reality," in Proceeding of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22-26 March 2020 (IEEE), 654–655. doi:10.1109/vrw50115.2020.00178

Lobachev, O., Berthold, M., Pfeffer, H., Guthe, M., and Steiniger, B. S. (2021). Inspection of Histological 3d Reconstructions in Virtual Reality. *Front. Virtual Real.* 2. doi:10.3389/frvir.2021.628449

Lohre, R., Warner, J. J. P., Athwal, G. S., and Goel, D. P. (2020). The Evolution of Virtual Reality in Shoulder and Elbow Surgery. *JSES Int.* 4, 215–223. doi:10.1016/j.jseint.2020.02.005

Manzo, C., and Garcia-Parajo, M. F. (2015). A Review of Progress in Single Particle Tracking: from Methods to Biophysical Insights. *Rep. Prog. Phys.* 78, 124601. doi:10.1088/0034-4885/78/12/124601

Martirosov, S., Hořejší, P., Kopeček, P., Bureš, M., and Šimon, M. (2021). The Effect of Training in Virtual Reality on the Precision of Hand Movements. *Appl. Sci.* 11, 8064. doi:10.3390/app11178064

Masson, J. B., Laurent, F., Cardona, A., Barré, C., Skatchkovsky, N., Zlatic, M., et al. (2020). Identifying Neural Substrates of Competitive Interactions and Sequence Transitions during Mechanosensory Responses in drosophila. *Plos Genet.* 16, e1008589. doi:10.1371/journal.pgen.1008589

Matthews, D. (2018). Virtual-reality Applications Give Science a New Dimension. *Nature* 557, 127–128. doi:10.1038/d41586-018-04997-2

Mostafa, Y. (2012). *Learning from Data : A Short Course (United States: AMLBook.Com)*.

Naviscope (2021). Image-guided Navigation and VIsualization of Large Data Sets in Live Cell Imaging and microSCOPy.

O'Connor, M., Deeks, H. M., Dawn, E., Metatla, O., Roudaut, A., Sutton, M., et al. (2018). Sampling Molecular Conformations and Dynamics in a Multiuser Virtual Reality Framework. *Sci. Adv.* 4, eaat2731. doi:10.1126/sciadv.aat2731

Ong, C. S., Krishnan, A., Huang, C. Y., Spevak, P., Vricella, L., Hibino, N., et al. (2018). Role of Virtual Reality in Congenital Heart Disease. *Congenit. Heart Dis.* 13, 357–361. doi:10.1111/chd.12587

Pastel, S., Chen, C.-H., Martin, L., Naujoks, M., Petri, K., and Witte, K. (2021). Comparison of Gaze Accuracy and Precision in Real-World and Virtual Reality. *Virtual Reality* 25, 175–189. doi:10.1007/s10055-020-00449-3

Patel, B. N., Rosenberg, L., Willcox, G., Baltaxe, D., Lyons, M., Irvin, J., et al. (2019). Human-machine Partnership with Artificial Intelligence for Chest Radiograph Diagnosis. *NPJ Digit Med.* 2, 111. doi:10.1038/s41746-019-0189-7

Pavez, E., Chou, P. A., de Queiroz, R. L., and Ortega, A. (2018). Dynamic Polygon Clouds: Representation and Compression for VR/AR. *APSIPA Trans. Signal Inf. Process.* 7, 1–22. doi:10.1017/atsip.2018.15

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Machine Learn. Res.* 12, 2825–2830.

Peng, H., Ruan, Z., Long, F., Simpson, J. H., and Myers, E. W. (2010). V3d Enables Real-Time 3d Visualization and Quantitative Analysis of Large-Scale Biological Image Data Sets. *Nat. Biotechnol.* 28, 348–353. doi:10.1038/nbt.1612

Pfandler, M., Lazarovici, M., Stefan, P., Wucherer, P., and Weigl, M. (2017). Virtual Reality-Based Simulators for Spine Surgery: a Systematic Review. *Spine J.* 17, 1352–1363. doi:10.1016/j.spinee.2017.05.016

Pfeiffer, M., Kenngott, H., Preukschas, A., Huber, M., Bettscheider, L., Müller-Stich, B., et al. (2018). IMHOTEP: Virtual Reality Framework for Surgical Applications. *Int. J. Comput. Assist. Radiol. Surg.* 13, 741–748. doi:10.1007/s11548-018-1730-x

Pinter, C., Lasso, A., Choueib, S., Asselin, M., Fillion-Robin, J. C., Vimort, J. B., et al. (2020). SlicerVR for Medical Intervention Training and Planning in Immersive Virtual Reality. *IEEE Trans. Med. Robot Bionics* 2, 108–117. doi:10.1109/tmrb.2020.2983199

PrecisionOS (2021). *Virtual Reality (VR) Surgery Training & Simulators*.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2016). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv preprint*. arXiv: 1706.02413.

Quero, G., Lapergola, A., Soler, L., Shahbaz, M., Hostettler, A., Collins, T., et al. (2019). Virtual and Augmented Reality in Oncologic Liver Surgery. *Surg. Oncol. Clin. N. Am.* 28, 31–44. doi:10.1016/j.soc.2018.08.002

Raimondi, F., Vida, V., Godard, C., Bertelli, F., Reffo, E., Boddaert, N., et al. (2021). Fast-track Virtual Reality for Cardiac Imaging in Congenital Heart Disease. *J. Card. Surg.* 36, 2598–2602. doi:10.1111/jocs.15508

Ramirez, P. Z., Paternesi, C., Luigi, L. D., Lella, L., Gregorio, D. D., and Stefano, L. D. (2020). "Shooting Labels: 3d Semantic Labeling by Virtual Reality," in Proceeding of the 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality, Utrecht, Netherlands, 14-18 Dec. 2020 (IEEE), 99–106. doi:10.1109/AIVR50618.2020.00027

Reitinger, B., Bornik, A., Beichel, R., and Schmalstieg, D. (2006). Liver Surgery Planning Using Virtual Reality. *IEEE Comput. Graph Appl.* 26, 36–47. doi:10.1109/mcg.2006.131

Ruiz, F., Neiva-Vaz, C., Vazquez, M.-P., Masson, J.-B., El Beheiry, M., Pannier, S., et al. (2021). Management of Ischiopagus Twin Separation with a Focus on W-S Incision Design. *J. Pediatr. Surg. Case Rep.* 64, 101747. doi:10.1016/j.epsc.2020.101747

Sadeghi, A. H., Bakhuis, W., Van Schaagen, F., Oei, F. B. S., Bekkers, J. A., Maat, A. P. W. M., et al. (2020). Immersive 3d Virtual Reality Imaging in Planning Minimally Invasive and Complex Adult Cardiac Surgery. *Eur. Heart J. - Digital Health* 1, 62–70. doi:10.1093/ehjdh/ztaa011

Safaryan, K., and Mehta, M. R. (2021). Enhanced Hippocampal Theta Rhythmicity and Emergence of Eta Oscillation in Virtual Reality. *Nat. Neurosci.* 24, 1065–1070. doi:10.1038/s41593-021-00871-z

Salvatore, S. D., Vadalà, G., Oggiano, L., Russo, F., Ambrosio, L., and Costici, P. F. (2021). Virtual Reality in Preoperative Planning of Adolescent Idiopathic Scoliosis Surgery Using Google Cardboard. *Neurospine* 18, 199–205. doi:10.14245/ns.2040574.287

Sellés, T. S. (2013). *GaiaVR*.

Seymour, N. E. (2008). Vr to or: A Review of the Evidence that Virtual Reality Simulation Improves Operating Room Performance. *World J. Surg.* 32, 182–188. doi:10.1007/s00268-007-9307-9

Shao, X., Yuan, Q., Qian, D., Ye, Z., Chen, G., le Zhuang, K., et al. (2020). Virtual Reality Technology for Teaching Neurosurgery of Skull Base Tumor. *BMC Med. Educ.* 20, 3. doi:10.1186/s12909-019-1911-5

Shattuck, D. W. (2018). Multiuser Virtual Reality Environment for Visualising Neuroimaging Data. *Healthc. Technol. Lett.* 5, 183–188. doi:10.1049/htl.2018.5077

Sommer, C., Straehle, C., Köthe, U., and Hamprecht, F. A. (2011). "Ilastik: Interactive Learning and Segmentation Toolkit," in Proceeding of the 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, 30 March-2 April 2011 (IEEE), 230–233. doi:10.1109/ISBI.2011.5872394

Spark, A., Kitching, A., Esteban-Ferrer, D., Handa, A., Carr, A. R., Needham, L. M., et al. (2020). vLUME: 3d Virtual Reality for Single-Molecule Localization Microscopy. *Nat. Methods* 17, 1097–1099. doi:10.1038/s41592-020-0962-1

Stefani, C., Lacy-Hulbert, A., and Skillman, T. (2018). Confocalvr: Immersive Visualization for Confocal Microscopy. *J. Mol. Biol.* 430, 4028–4035. doi:10.1016/j.jmb.2018.06.035

Stets, J. D., Sun, Y., Corning, W., and Greenwald, S. W. (2017). Visualization and Labeling of point Clouds in Virtual Reality. *SIGGRAPH Asia 2017 Posters*. New York, NY: Association for Computing Machinery, 2. doi:10.1145/3145690.3145729

Sun, G. J., Sailor, K. A., Mahmood, Q. A., Chavali, N., Christian, K. M., Song, H., et al. (2013). Seamless Reconstruction of Intact Adult-Born Neurons by Serial End-Block Imaging Reveals Complex Axonal Guidance and Development in the Adult hippocampus. *J. Neurosci.* 33, 11400–11411. doi:10.1523/JNEUROSCI.1374-13.2013

SurgicalTheater (2021). *Experience 360°XR ANATOMY*.

syGlass (2021). *Scientific Virtual Reality*.

Theart, R. P., Loos, B., and Niesler, T. R. (2017). Virtual Reality Assisted Microscopy Data Visualization and Colocalization Analysis. *BMC Bioinformatics* 18, 64. doi:10.1186/s12859-016-1446-2

Thomsen, A. S., Bach-Holm, D., Kjærbo, H., Højgaard-Olsen, K., Subhi, Y., Saleh, G. M., et al. (2017). Operating Room Performance Improves after Proficiency-Based Virtual Reality Cataract Surgery Training. *Ophthalmology* 124, 524–531. doi:10.1016/j.ophtha.2016.11.015

Tomikawa, M., Hong, J., Shiotani, S., Tokunaga, E., Konishi, K., Ieiri, S., et al. (2010). Real-time 3-dimensional Virtual Reality Navigation System with Open MRI for Breast-Conserving Surgery. *J. Am. Coll. Surg.* 210, 927–933. doi:10.1016/j.jamcollsurg.2010.01.032

UnityTechnologies (2021). *Unity*.

Uppot, R. N., Laguna, B., McCarthy, C. J., De Novi, G., Phelps, A., Siegel, E., et al. (2019). Implementing Virtual and Augmented Reality Tools for Radiology Education and Training, Communication, and Clinical Care. *Radiology* 291, 570–580. doi:10.1148/radiol.2019182210

Uruthiralingam, U., and Rea, P. M. (2020). Augmented and Virtual Reality in Anatomical Education - A Systematic Review. *Adv. Exp. Med. Biol.* 1235, 89–101. doi:10.1007/978-3-030-37639-0_5

Usher, W., Klacansky, P., Federer, F., Bremer, P. T., Knoll, A., Yarch, J., et al. (2017). A Virtual Reality Visualization Tool for Neuron Tracing. *IEEE Trans. Vis. Comput. Graph* 24, 994–1003. doi:10.1109/TVCG.2017.2744079

Venkatesan, M., Mohan, H., Ryan, J. R., Schürch, C. M., Nolan, G. P., Frakes, D. H., et al. (2021). Virtual and Augmented Reality for Biomedical Applications. *Cell Rep Med* 2, 100348. doi:10.1016/j.xcrm.2021.100348

Verhey, J. T., Haglin, J. M., Verhey, E. M., and Hartigan, D. E. (2020). Virtual, Augmented, and Mixed Reality Applications in Orthopedic Surgery. *Int. J. Med. Robot* 16, e2067. doi:10.1002/rcs.2067

Wake, N., Nussbaum, J. E., Elias, M. I., Nikas, C. V., and Bjurlin, M. A. (2020). 3d Printing, Augmented Reality, and Virtual Reality for the Assessment and Management of Kidney and Prostate Cancer: A Systematic Review. *Urology* 143, 20–32. doi:10.1016/j.urology.2020.03.066

Walbron, P., Thomazeau, H., and Sirveaux, F. (2019). "Virtual reality simulation" in der Orthopädie und Unfallchirurgie in Frankreich. *Unfallchirurg* 122, 439–443. doi:10.1007/s00113-019-0649-1

Wang, S. S., Zhang, S. M., and Jing, J. J. (2012). Stereoscopic Virtual Reality Models for Planning Tumor Resection in the Sellar Region. *BMC Neurol.* 12, 146. doi:10.1186/1471-2377-12-146

Wang, Y., Li, Q., Liu, L., Zhou, Z., Ruan, Z., Kong, L., et al. (2019). Teravr Empowers Precise Reconstruction of Complete 3-d Neuronal Morphology in the Whole Brain. *Nat. Commun.* 10, 3474. doi:10.1038/s41467-019-11443-y

Wirth, F., Quehl, J., Ota, J., and Stiller, C. (2019). "PointAtMe: Efficient 3d point Cloud Labeling in Virtual Reality," in Proceeding of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9-12 June 2019 (IEEE), 1693–1698. doi:10.1109/ivs.2019.8814115

Wisotzky, E. L., Rosenthal, J.-C., Eisert, P., Hilsmann, A., Schmid, F., Bauer, M., et al. (2019). "Interactive and Multimodal-Based Augmented Reality for Remote Assistance Using a Digital Surgical Microscope," in Proceeding of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23-27 March 2019 (IEEE), 1477–1484. doi:10.1109/vr.2019.8797682

Yoo, J. S., Patel, D. S., Hrynewycz, N. M., Brundage, T. S., and Singh, K. (2019). The Utility of Virtual Reality and Augmented Reality in Spine Surgery. *Ann. Transl Med.* 7, S171. doi:10.21037/atm.2019.06.38

Zheng, Z., Lauritzen, J. S., Perlman, E., Robinson, C. G., Nichols, M., Milkie, D., et al. (2018). A Complete Electron Microscopy Volume of the Brain of Adult drosophila Melanogaster. *Cell* 174, 730–e22. doi:10.1016/j.cell.2018.06.019

Conflict of Interest: MB and J-BM are cofounders, shareholders and, respectively, Chief Technology Officer (CTO) and Chief Scientific Officer (CSO) of AVATAR MEDICAL SAS, a startup that commercializes software for surgery planning in virtual reality. The DIVA software used in this study is not being commercialized by AVATAR MEDICAL SAS also the company's technology is based on the same technology. The DIVA software which serves as base for this study is freely available and is reported in El Beheiry, et al. (2020). All developments within this paper are open source. HV was employed by Sanofi R&D.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

1 INTRODUCTION

This supplementary shows examples of various situations, amongst them cases where the data was complex, the tagging imperfect, or the learning not sufficient to produce acceptable results. We focused some examples on complex medical data where the current health state of the patient induced degraded images such as the one in fig S5. We also provided video captures of several processes to show how the pipeline is used, how the various interfaces work, and the advantages of VR for visualization and annotation.

The entire pipeline for *Voxel Learning* is accessible in `https://github.com/DecBayComp/VoxelLearning`. It includes an executable of the software DIVA, as well as DIVA Cloud source code and thorough instructions for its installation. In order to allow reproducibility, we provide in the directory `https://github.com/DecBayComp/VoxelLearning/materials/data_examples/` all the examples presented in the main and supplementary data. For each image, an already adjusted transfer function, tags performed in VR and trained classifiers are available in JSON format. Any user can then load and use them to recapitulate the results presented throughout this article.

2 EXAMPLES OF APPLICATION

2.1 Medical images

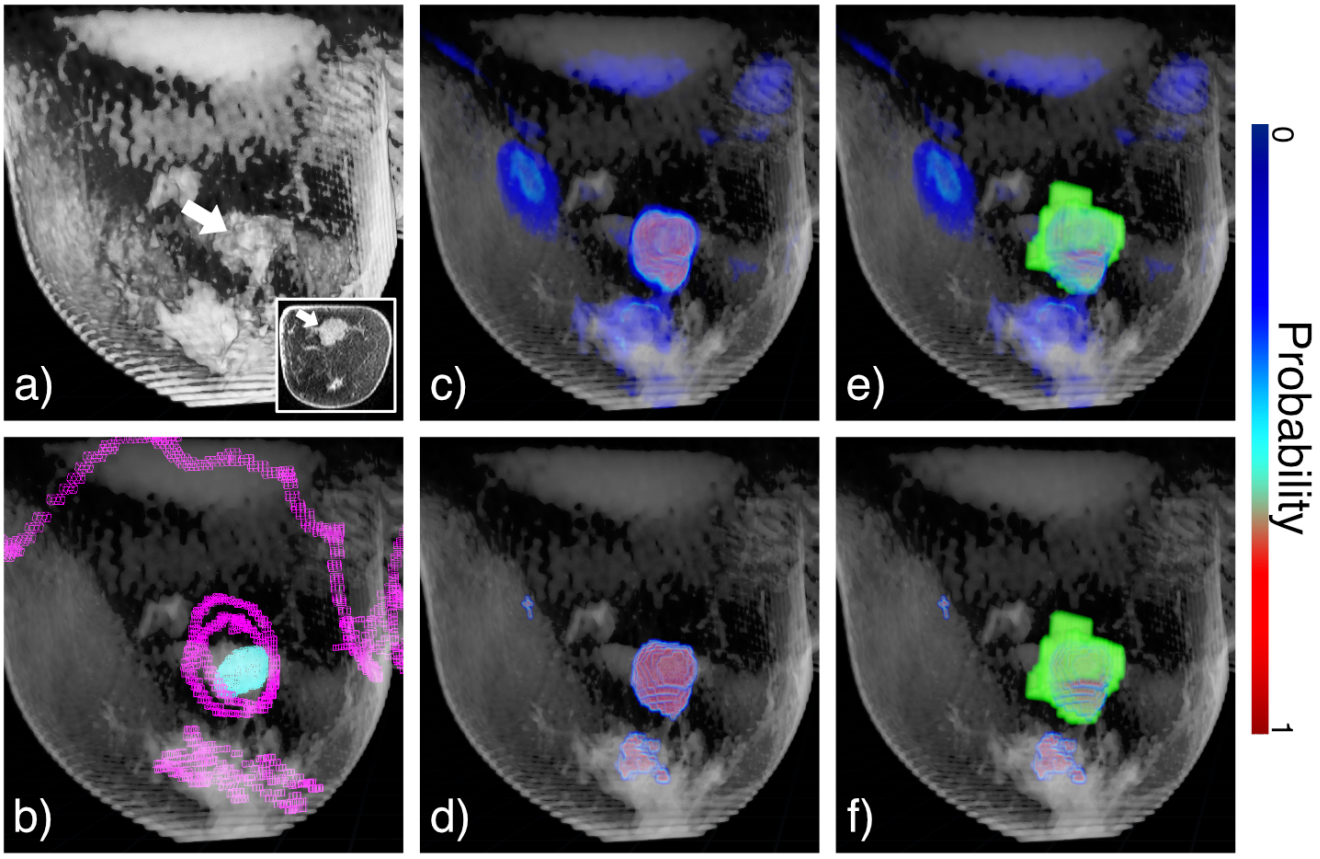


Figure S1. Annotation in DIVA on MRI of breast tumor (white arrow) from the Cancer Imaging Archive (TCIA) from the RIDER Breast MRI collection subject RIDER-1627409910.**a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

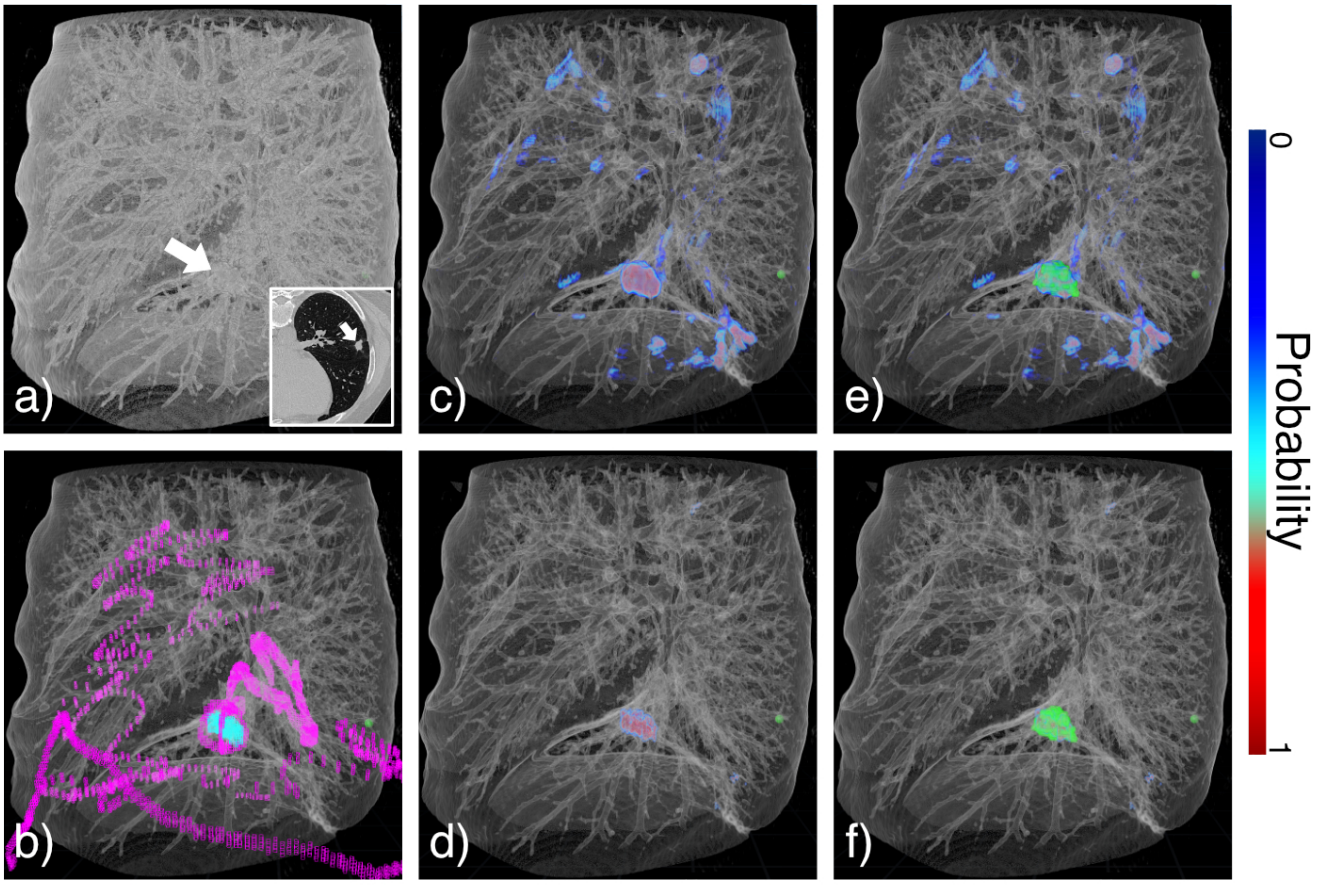


Figure S2. Annotation in DIVA on CT-scan of lung tumor (white arrow) from the Medical Segmentation Decathlon (MSD) challenge : lung_003. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

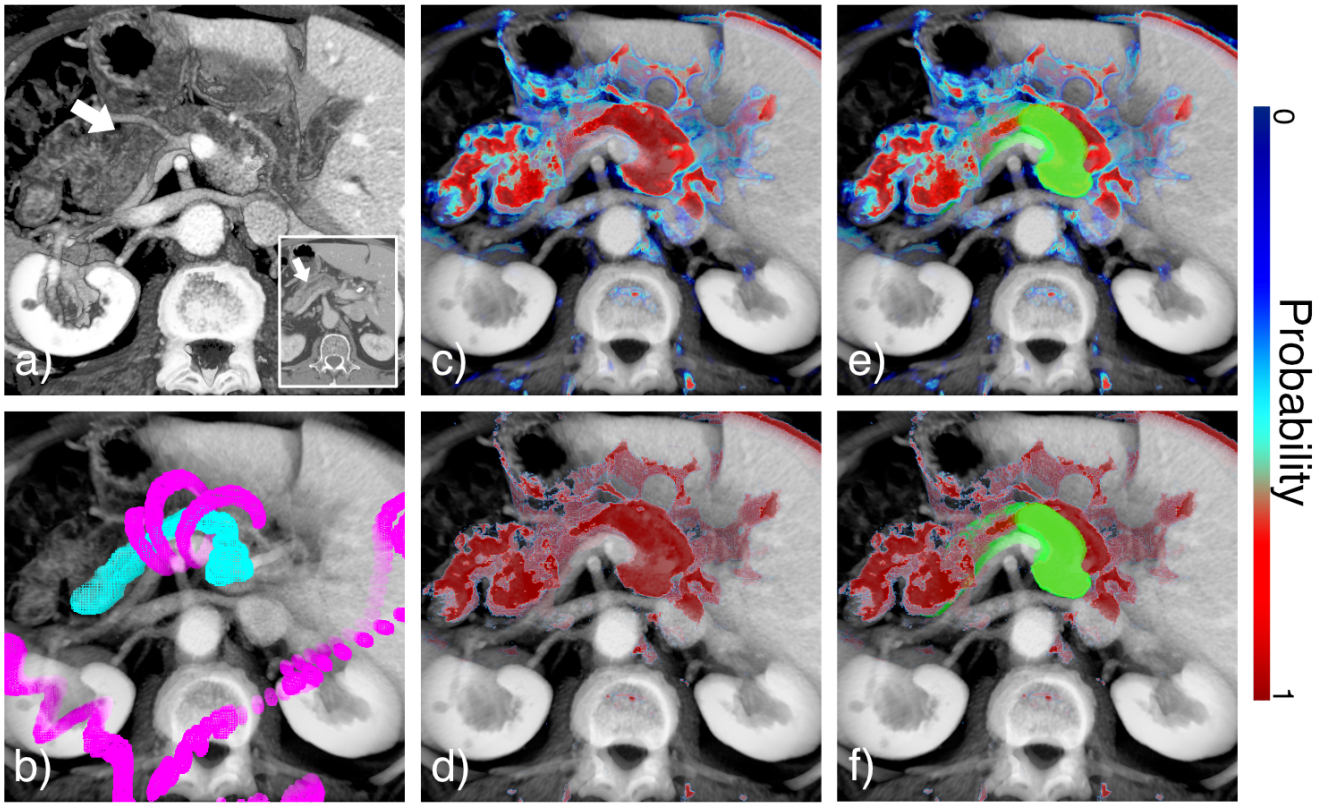


Figure S3. Annotation in DIVA on CT-scan of pancreas (white arrow) from the Medical Segmentation Decathlon (MSD) challenge : pancreas_001. **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

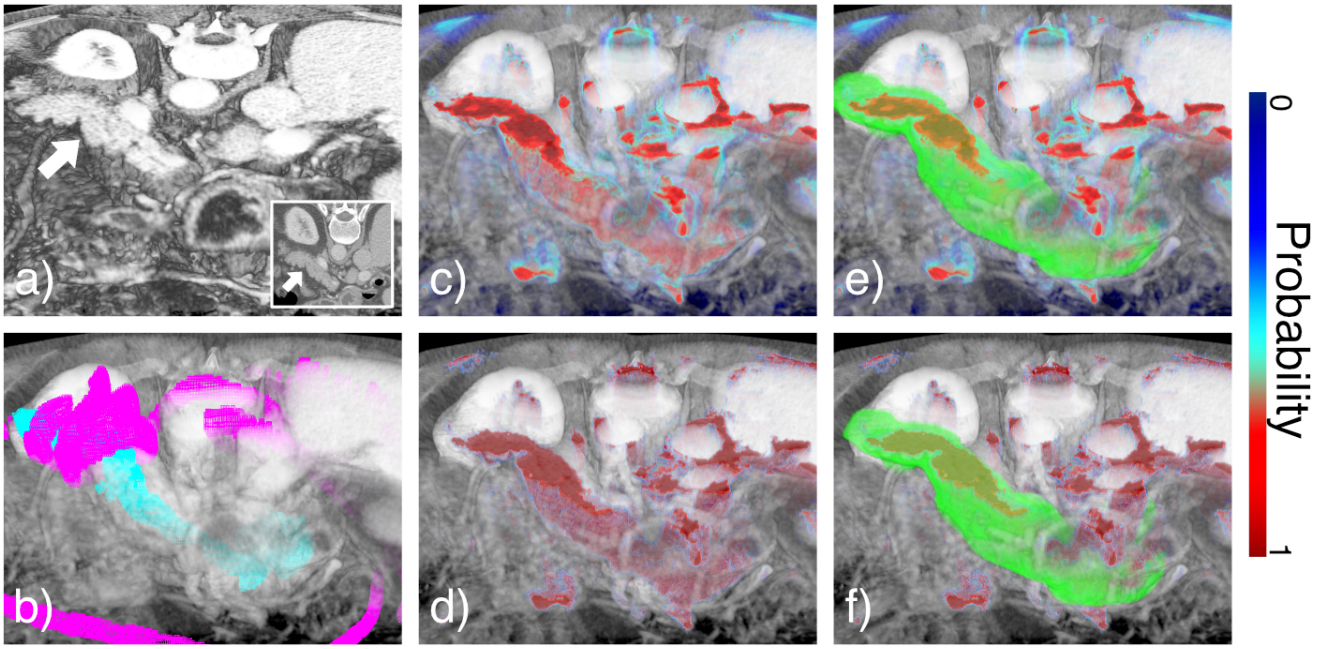


Figure S4. Annotation in DIVA on CT-scan of pancreas (white arrow) from the Medical Segmentation Decathlon (MSD) challenge : pancreas_004. **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

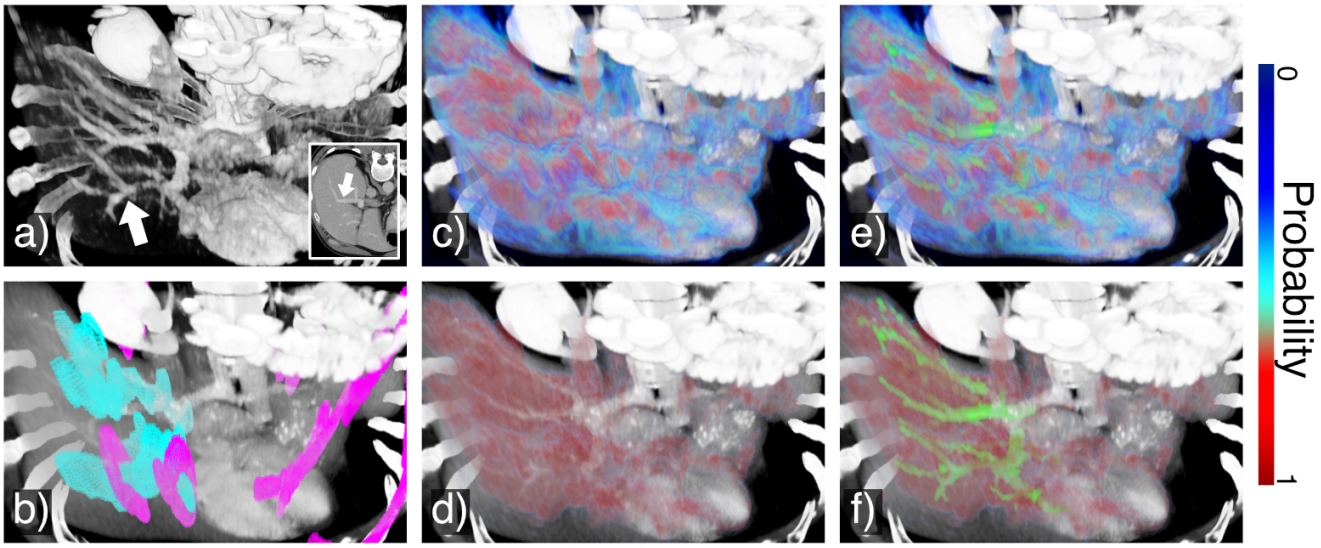


Figure S5. Annotation in DIVA on CT-scan of hepatic vessel (white arrow) from the Medical Segmentation Decathlon (MSD) challenge : hepaticvessel_001. **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

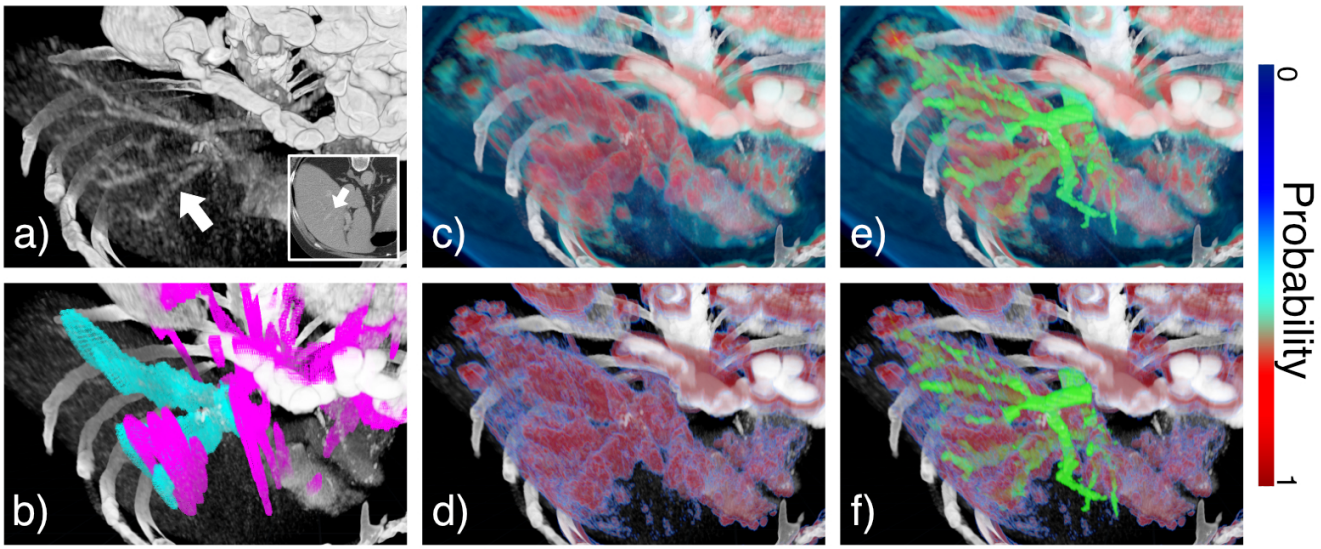


Figure S6. Annotation in DIVA on CT-scan of hepatic vessel (white arrow) from the Medical Segmentation Decathlon (MSD) challenge : hepaticvessel_002. **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

2.2 Microscopic images

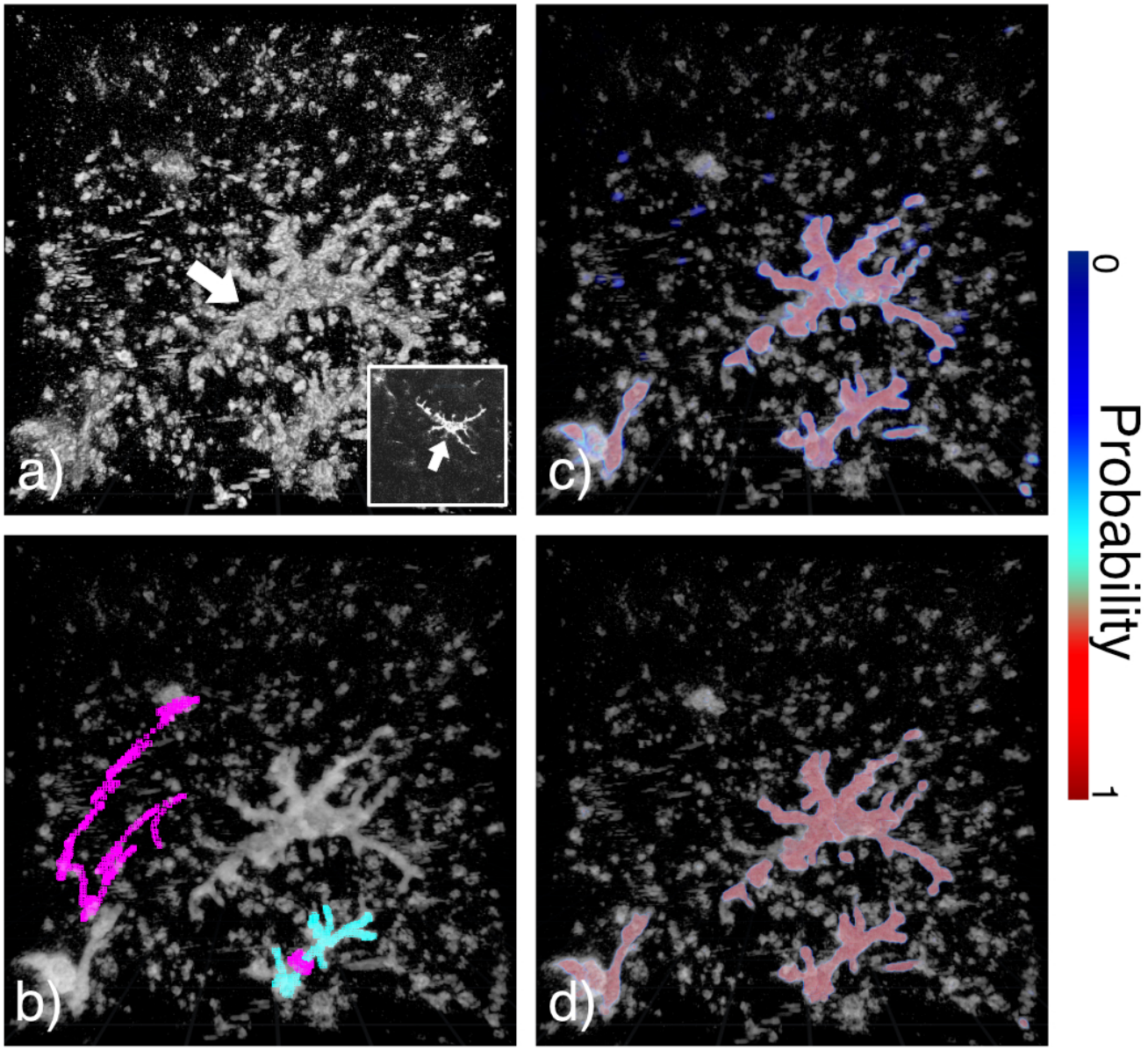


Figure S7. Annotation in DIVA on two-photon microscopy images of mouse microglia (white arrow), courtesy of Kurt Sailor (Institut Pasteur). **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

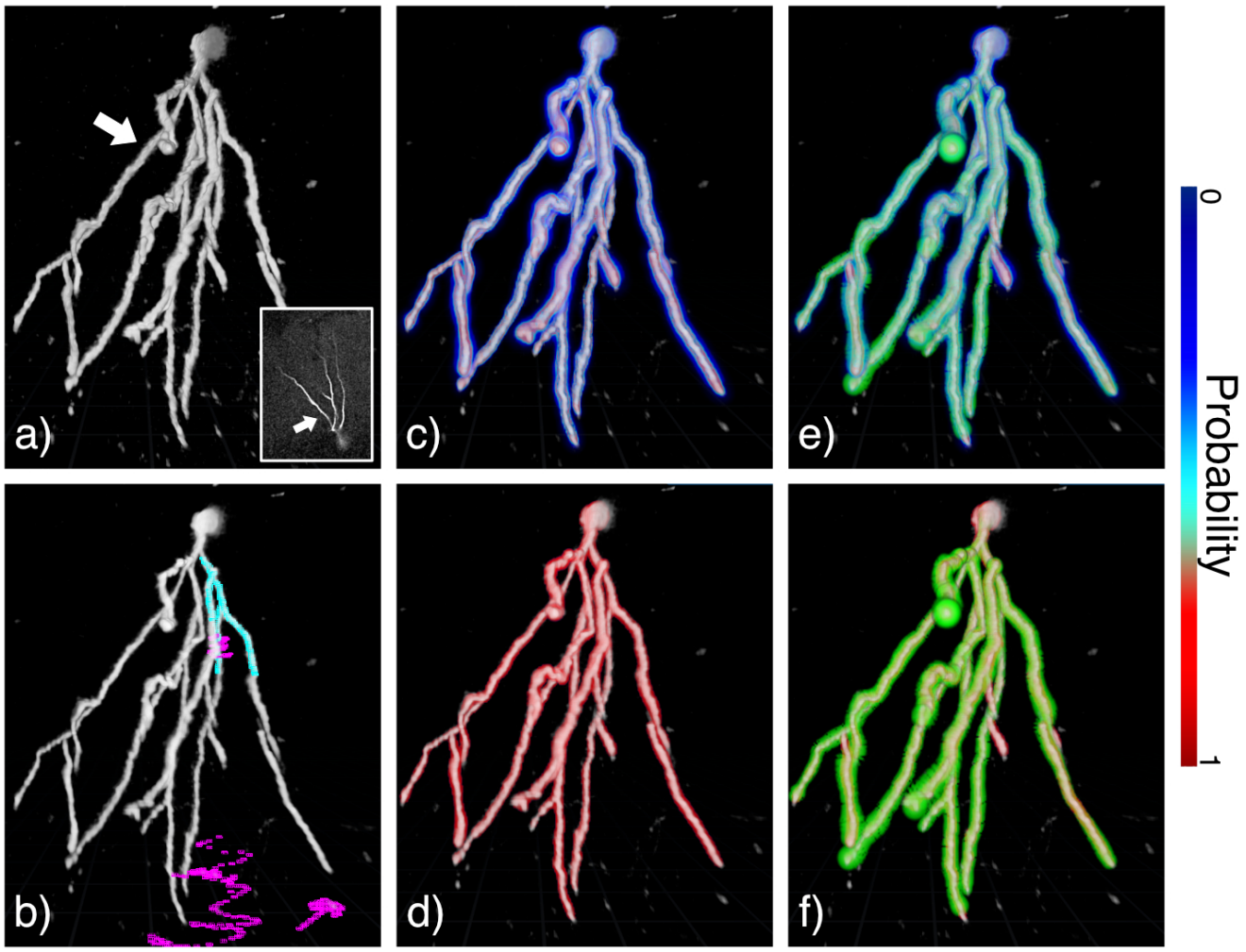


Figure S8. Annotation in DIVA on SEBI images of a mouse hippocampus neuron (white arrow), courtesy of Kurt Sailor (Institut Pasteur). **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. **b)** Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. **c-d)** Overlay of raw data in gray and output probabilities respectively for RFC and *strong learner*. Colorscale for probabilities is indicated at the right.

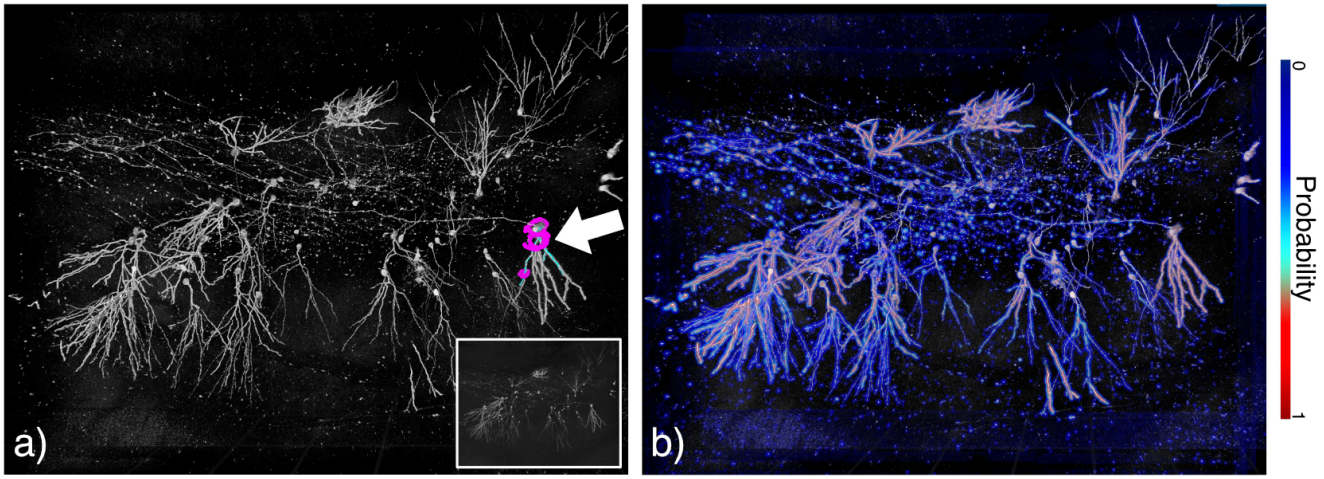


Figure S9. Annotation in DIVA on full SEBI images of mouse hippocampus neurons, courtesy of Kurt Sailor (Institut Pasteur). **a)** Raw data visualized in 3D on DIVA and as a z-stack in the bottom right corner. Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta. Tags are indicated with a white arrow. **b)** Overlay of raw data in gray and output probabilities for RFC. Colorscale for probabilities is indicated at the right.

3 FEATURES IMPORTANCE

We show here two examples of feature importance evaluation for two medical image stacks and one microscopy stack.

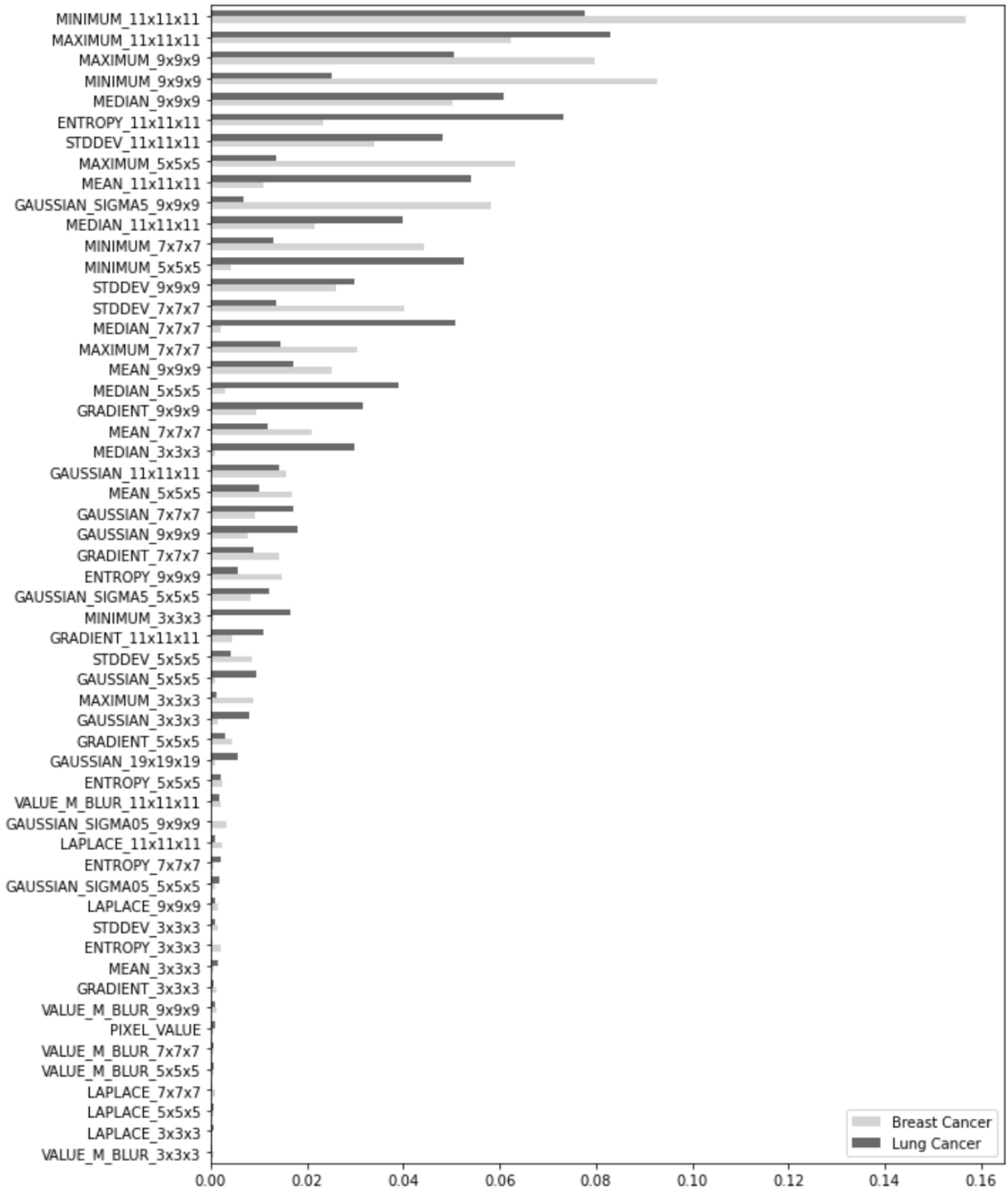


Figure S10. Features importance for RFC trained on examples of MRI of breast cancer and CT-scan of lung cancer. The comprehensive list of 56 features is here available.

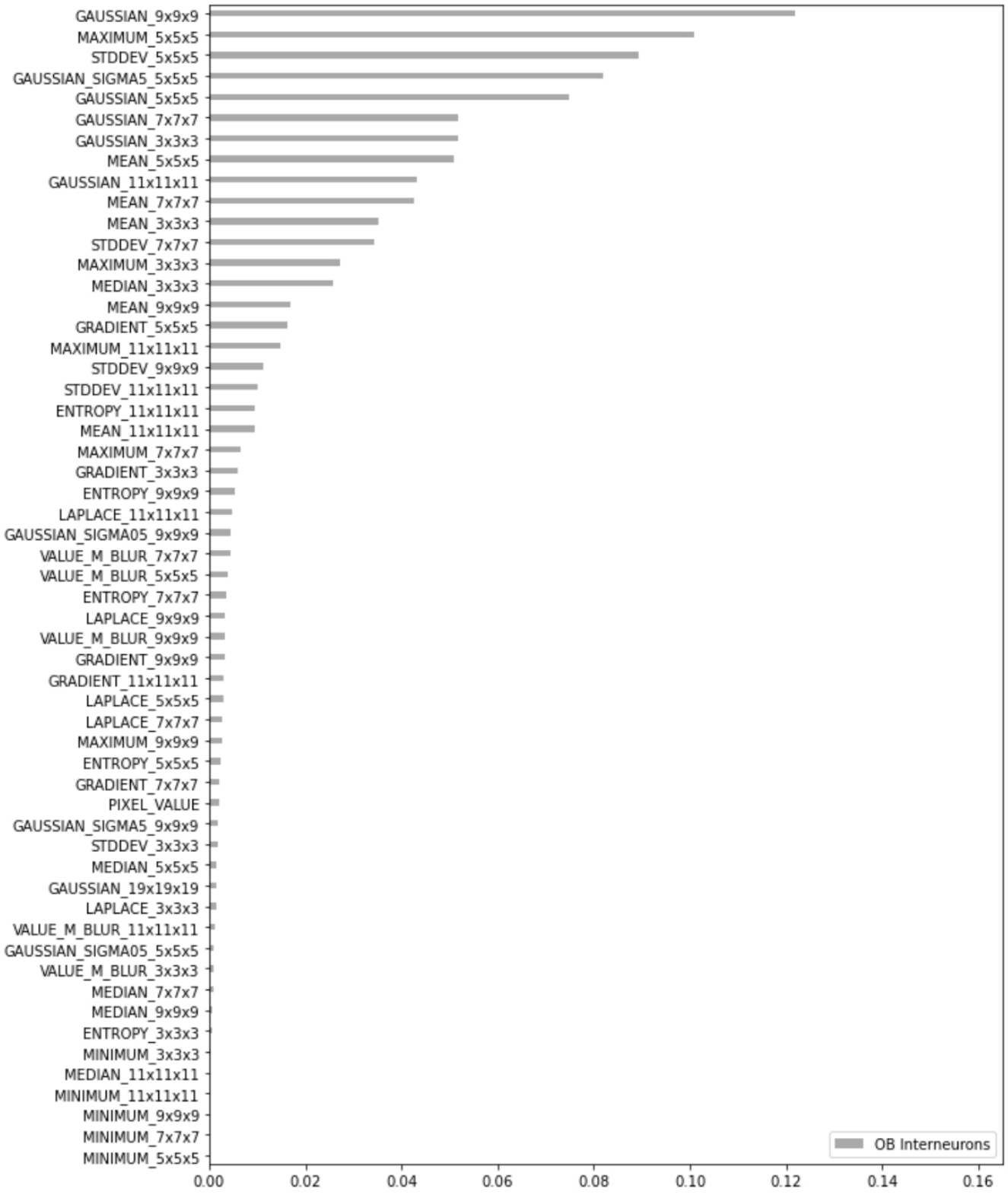


Figure S11. Features importance for RFC trained on mouse olfactory bulb (OB) interneurons example.

4 ANNOTATING IN 3D VS ANNOTATING IN 2D

We show in this section an example of a comparison between our approach to accelerated annotation and ilastik interactive segmentation (*Sommer et al., 2011*). Figure S12 shows the benefit of VR and transfer function adjustment in the visualization of complex 3D structures, here on an annotation task of hepatic vessels. In such examples, structures of interest are difficult to see in 2D sections. Volumetric reconstruction in 3D allows efficient rendering of the structure, improving thus the tagging experience. 3D motion accelerates the annotation by removing the necessity to scroll from one slide to another.

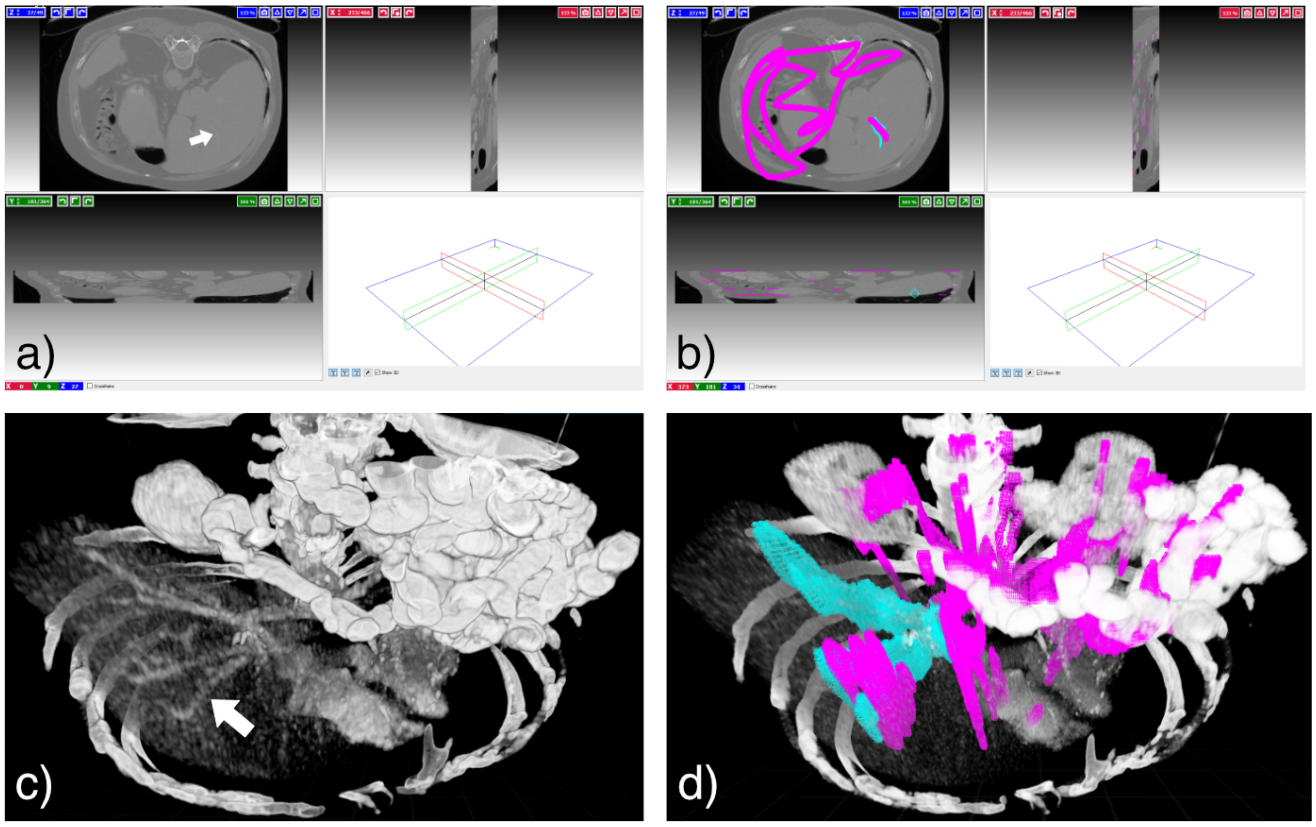


Figure S12. User experience comparison in visualization of an example CT-scan of hepatic vessel (white arrow in panels **a)** and **c)**) from the Medical Segmentation Decathlon (MSD) challenge : hepaticvessel_002. **a)** and **b)** Screenshots of ilastik interface with planar sections in the three natural axes. **c)** and **d)** Screenshots of DIVA interface, after appropriate transfer function setting. Tags are shown in **b)** and **d)** in cyan for the structure of interest, here hepatic vessels, and in magenta for the background.

We show in Figure S13 a performance comparison between ilastik and Voxel Learning applied on medical example images. Our tests were conducted in a one-shot paradigm, where the user provides limited tags: one line in the structure of interest, one line out of it. The tagging procedure should only last a few seconds. When compared to expert segmentation, we show on medical examples that our technique yields higher Dice coefficients than ilastik. Importing tags done in VR into ilastik improves its performance, while not reaching similar scores as Voxel Learning. Note that this comparison does not involve the quality

of learning as we use similar approaches as ilastik. It rather shows the expected gain from the use of VR annotation in order to provide a one-shot segmentation of the data.

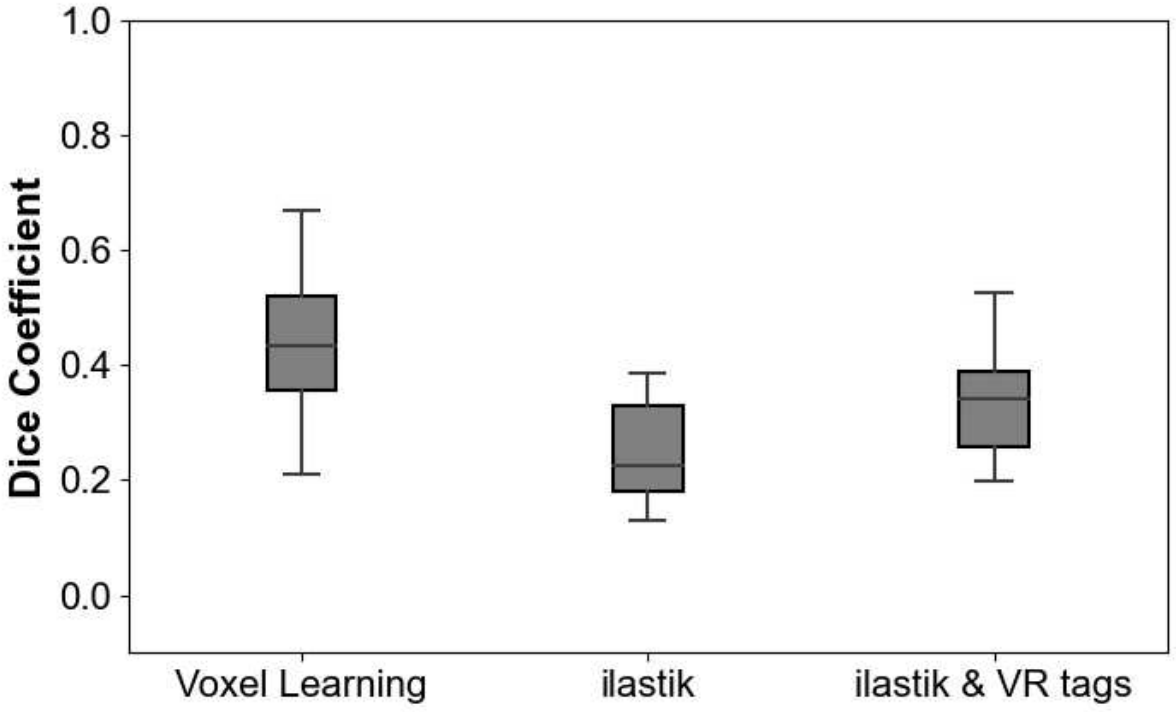
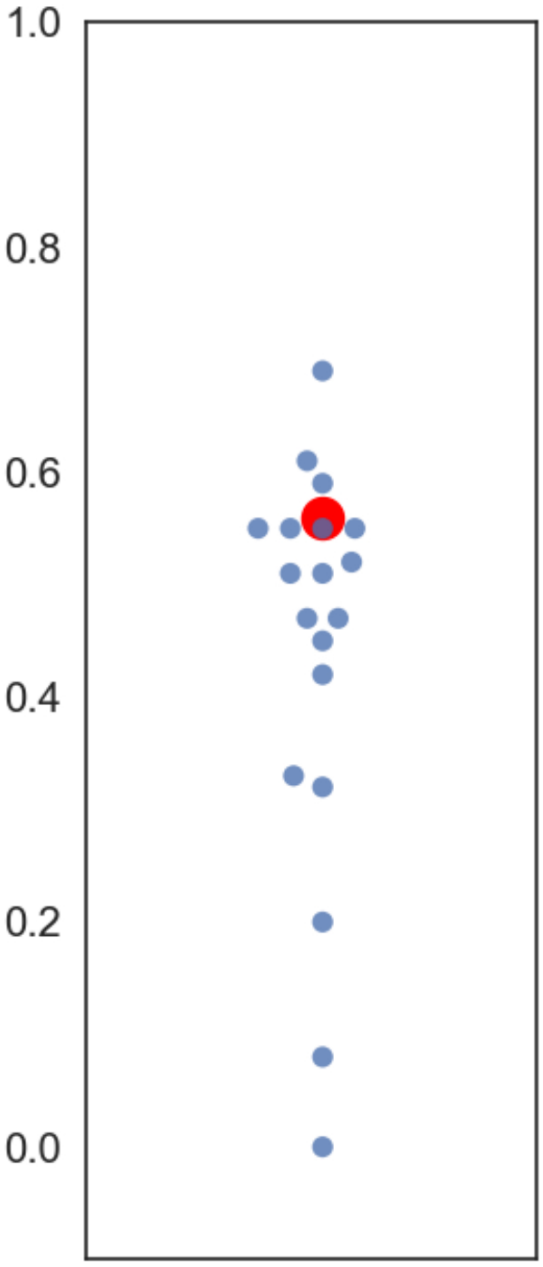


Figure S13. Performance comparison between Voxel Learning and ilastik on the 8 medical examples used throughout this article: two occurences of breast tumor MRI, and CT-scan of lung tumor, hepatic vessel and pancreas. Few tags are performed in VR for Voxel Learning, and in a 2D section in ilastik. The distribution of Dice coefficient over the different examples is shown here for Voxel Learning annotation using VR tags, and ilastik annotations respectively using 2D and VR tags.

5 A QUANTIFICATION OF EFFICIENCY IN LEARNING

In order to provide a quantitative assessment of the quality of the annotation performed within the software, we compared it to segmentation pipelines on examples of CT-scan of lung tumor. Results are shown in Figure S14. Note that such algorithms are usually trained on vast databases already annotated by experts, while our approach is based on one-shot learning, thus requiring no previous knowledge. As these results may be corrected with additional tagging, our procedure can then be iterated and performance enhanced.

a)



b)

Team	Dice
nnU-Net	0.69
K.A.V.athlon	0.61
CerebriuDIKU	0.59
Voxel Learning	**0.56**
LS Wang's Group	0.55
MIMI	0.55
Lupin	0.55
SIAT_MIDS	0.55
NVDLMED	0.52
Whale	0.51
UBIlearn	0.51
VST	0.47
LfB	0.47
A-REUMI01	0.45
BCVuniandes	0.42
BUT	0.33
Lesswire1	0.32
AI-Med	0.2
EdwardMa12593	0.08
RegionTec	0

Figure S14. Performance comparison of different segmentation pipelines on examples of CT-scan of lung tumor. We used cropped images here for practicality, and performed a single round of tagging in VR, training and inference using RFC. Dice coefficient was then assessed for 32 different examples from the MSD, and compared with the results presented in Figure 3 of *Isensee et al, 2021*. Our method obtained an average Dice coefficient of 0.56 (denoted in red), ranking 4 out of 20 segmentation algorithms (denoted in blue). A graphical representation of this comparison can be seen in **a)** with supporting data in **b)**.

6 STATISTICAL RESULTS

Modality	File Name	Original Size (number of pixels)	VR Tagging Time (s)	Tags + cyan (number of pixels)	Tags - magenta (number of pixels)	Training Features Evaluation (s)		RFC	XGB	SVM	NBC	MLP	Strong
MRI *breast_cancer*	image_01	288x288x60	30	1180	4530	26	Training Time (s)	0.58	3.27	0.18	0.06	3.10	26.60
							Inference Time (s)	131.80	679.30	68.36	92.90	73.64	832.20
							Total Time (s)	188.38	738.57	124.54	148.96	132.74	914.80
							DICE	0.78	0.72	0.22	0.74	0.47	0.74
							RMSE	0.02	0.02	0.05	0.02	0.04	0.02
	image_02	288x288x60	30	1195	3102	14	Training Time (s)	0.60	3.10	0.21	0.35	3.58	33.30
							Inference Time (s)	58.00	286.00	29.30	39.90	31.30	365.00
							Total Time (s)	102.60	333.10	73.51	84.25	78.88	442.30
							DICE	0.61	0.62	0.22	0.43	0.50	0.60
							RMSE	0.04	0.04	0.12	0.07	0.05	0.04
CT *lung_cancer*	image_01	512x512x304	60	718	15793	165	Training Time (s)	2.15	8.94	0.93	0.12	9.06	57.57
							Inference Time (s)	725.00	3677.00	393.00	506.00	417.00	4390.00
							Total Time (s)	952.15	3910.94	618.93	731.12	651.06	4672.57
							DICE	0.59	0.25	0.03	0.02	0.00	0.35
							RMSE	0.04	0.04	0.07	0.16	0.52	0.01
	image_02	512x512x288	120	556	12003	68	Training Time (s)	1.44	10.40	0.53	0.08	5.23	43.40
							Inference Time (s)	297.00	1471.00	155.00	198.00	157.00	1675.00
							Total Time (s)	486.44	1669.40	343.53	386.08	350.23	1906.40
							DICE	0.76	0.58	0.00	0.00	0.03	0.62
							RMSE	0.03	0.03	0.29	0.28	0.11	0.01
CT *hepatic_vessel*	image_01	512x512x49	85	16721	30425	60	Training Time (s)	5.66	26.06	7.76	0.28	27.95	126.60
							Inference Time (s)	486.60	2308.00	232.20	311.20	242.20	3436.00
							Total Time (s)	637.26	2479.06	384.96	456.48	415.15	3707.60
							DICE	0.16	0.15	0.03	0.08	0.10	0.07
							RMSE	0.14	0.13	0.74	0.47	0.42	0.13
	image_02	512x512x66	120	16819	62660	43	Training Time (s)	5.80	34.00	15.20	0.51	46.60	166.00
							Inference Time (s)	285.00	1464.00	144.00	189.00	153.00	1992.00
							Total Time (s)	453.80	1661.00	322.20	352.51	362.60	2321.00
							DICE	0.30	0.19	0.02	0.07	0.08	0.11
							RMSE	0.25	0.41	0.58	0.29	0.26	0.18
CT *pancreas*	image_01	512x419x51	120	43541	114081	52	Training Time (s)	28.14	71.39	40.35	0.98	30.49	353.00
							Inference Time (s)	387.00	1877.00	178.00	237.00	188.00	2711.00
							Total Time (s)	587.14	2120.39	390.35	409.98	390.49	3236.00
							DICE	0.59	0.43	0.02	0.06	0.07	0.33
							RMSE	0.14	0.15	0.75	0.47	0.43	0.15
	image_02	512x512x107	120	18883	77333	60	Training Time (s)	21.80	63.10	15.10	5.70	90.20	262.00
							Inference Time (s)	343.00	1470.00	152.00	213.00	157.00	2136.00
							Total Time (s)	544.80	1713.10	347.10	398.70	427.20	2578.00
							DICE	0.39	0.23	0.06	0.06	0.08	0.15
							RMSE	0.19	0.22	0.48	0.49	0.32	0.24

Table S1. Different measurements collected when performing data annotation with DIVA on various medical examples. For each example, the file name and the size in pixels are available. We took track as well of the tagging time in VR, the number of positively tagged voxels (”+” in cyan) and negatively tagged voxels (”-” in magenta), and the time to calculate features for the training set. For each learner, we collected the training time, the inference time and the total time for the procedure. To assess the quality of annotation, we computed the Dice coefficient and the Residual Mean Square Error (RMSE) between our results and an expert segmentation. All time measurements are expressed in seconds.

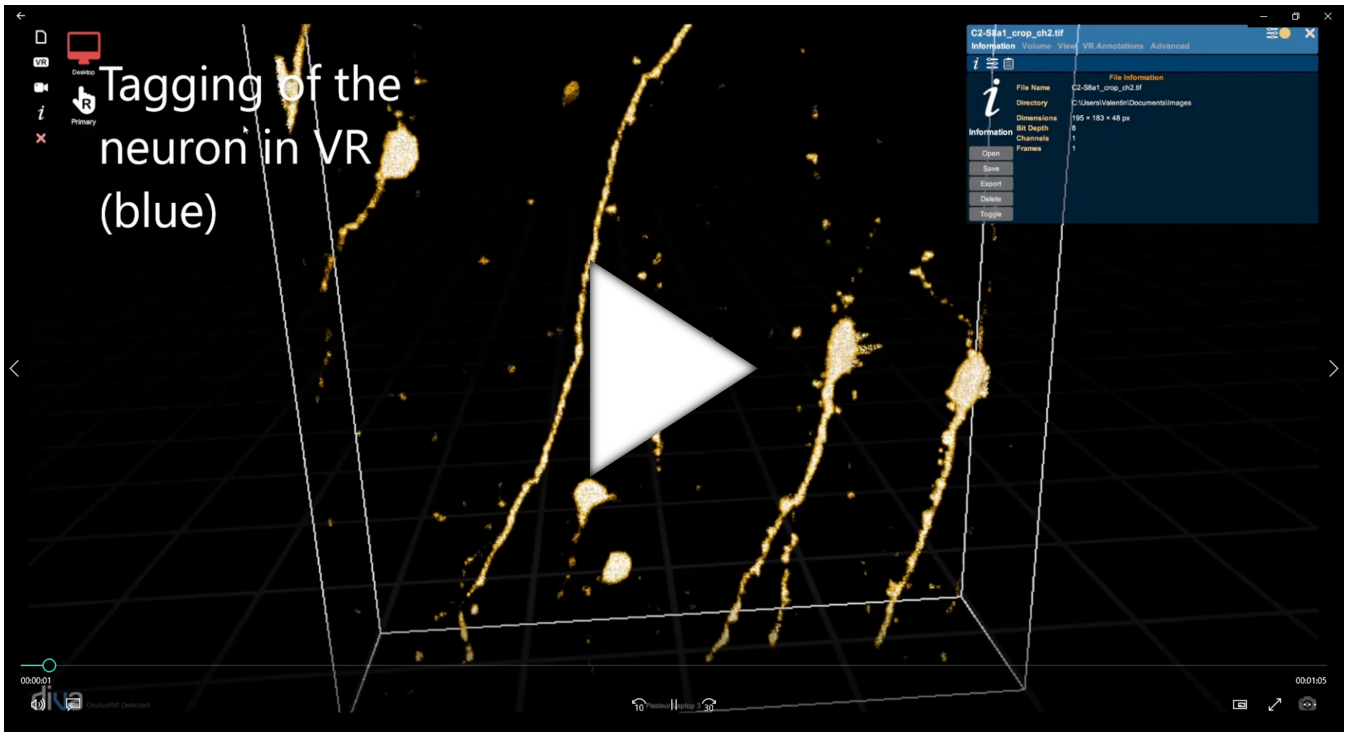
Modality	File Name	Original Size *(number of pixels)*	VR Tagging Time *(s)*	Tags + cyan *(number of pixels)*	Tags - magenta *(number of pixels)*	Training Features Evaluation *(s)*		RFC	XGB	SVM	NBC	MLP	Strong
Confocal ***interneurons_OB***	image_01	315x204x39	90	1008	2207	12	**Training Time** *(s)*	0.60	2.70	0.40	0.20	1.70	27.00
							Inference Time *(s)*	109.00	616.00	51.60	69.70	54.00	750.00
							Total Time (s)	211.60	720.70	154.00	171.90	157.70	879.00
Two-photon ***microglia***	image_01	255x255x63	60	1540	2914	30	**Training Time** *(s)*	0.70	3.60	0.40	0.28	3.80	36.00
							Inference Time *(s)*	224.00	555.00	178.00	75.00	58.00	654.00
							Total Time (s)	314.70	648.60	268.40	165.28	151.80	780.00
SEBI ***hippocampus***	image_01	214x296x125	90	720	957	55	**Training Time** *(s)*	0.30	2.60	0.20	0.15	0.34	17.00
							Inference Time *(s)*	218.00	1221.00	132.00	191.00	150.00	1283.00
							Total Time (s)	363.30	1368.60	277.20	336.15	295.34	1445.00
	image_02	1445x1127x263	80	2253	47032	2670	**Training Time** *(s)*	3.98	22.30	3.66	0.33	28.80	100.70
							Inference Time *(s)*	12410.00	52794.00	7742.00	9261.00	7591.00	75626.00
							Total Time (s)	15163.98	55566.30	10495.66	12011.33	10369.80	78476.70

Table S2. Different measurements collected when performing data annotation with DIVA on various neuronal microscopy examples. For each example, the file name and the size in pixels are available. We took track as well of the tagging time in VR, the number of positively tagged voxels ("+" in cyan) and negatively tagged voxels ("-" in magenta), and the time to calculate features for the training set. For each learner, we collected the training time, the inference time and the total time for the procedure. All time measurements are expressed in seconds.

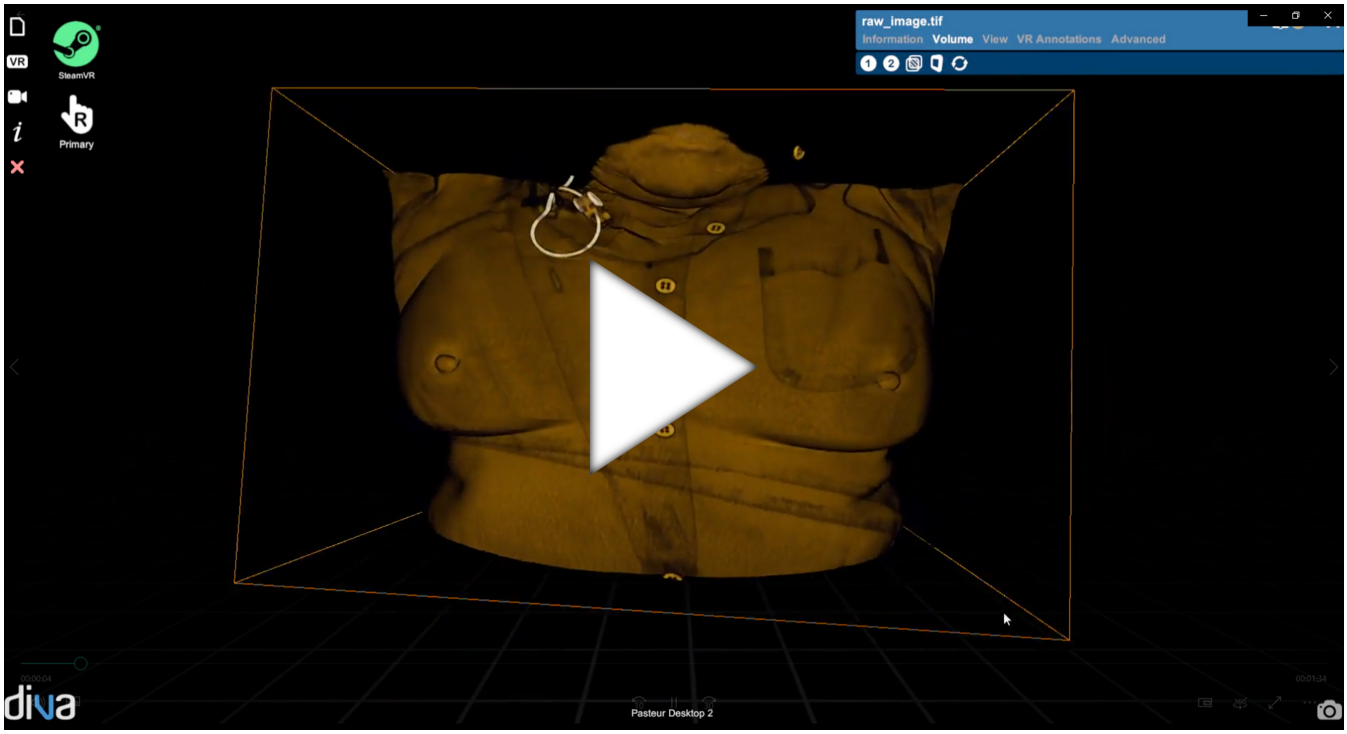
Modality	File Name	Image Size *(number of pixels)*	Desktop Frame Rate	VR Frame Rate
MRI *breast_cancer*	image_01	288x288x60	1100	51
	image_02	288x288x60	450	51
CT *lung_cancer*	image_01	512x512x304	100	16
	image_02	512x512x288	42	11
CT *hepatic_vessel*	image_01	512x512x49	350	50
	image_02	512x512x66	330	51
CT *pancreas*	image_01	512x419x51	500	110
	image_02	512x512x107	170	33
Confocal *interneurons_OB*	image_01	315x204x39	1050	115
Two-photon *microglia*	image_01	255x255x63	1000	51
SEBI *hippocampus*	image_01	214x296x125	330	33
	image_02	1445x1127x263	36	13

Table S3. Frame rate of the DIVA software for each image with a sampling resolution of 73 (parameter in DIVA from 0 to 100). The *Desktop Frame Rate* corresponds to the frame rate when using the Desktop interface (3D visualization of the volume on a 2D monitor). The *VR Frame Rate* corresponds to the frame rate when viewing and manipulating the volume in the virtual environment. These values are context-dependent and strongly vary upon user's actions in VR. If he navigates through the volume with the VR headset or moves too quickly, the frame rate may indeed be impacted.

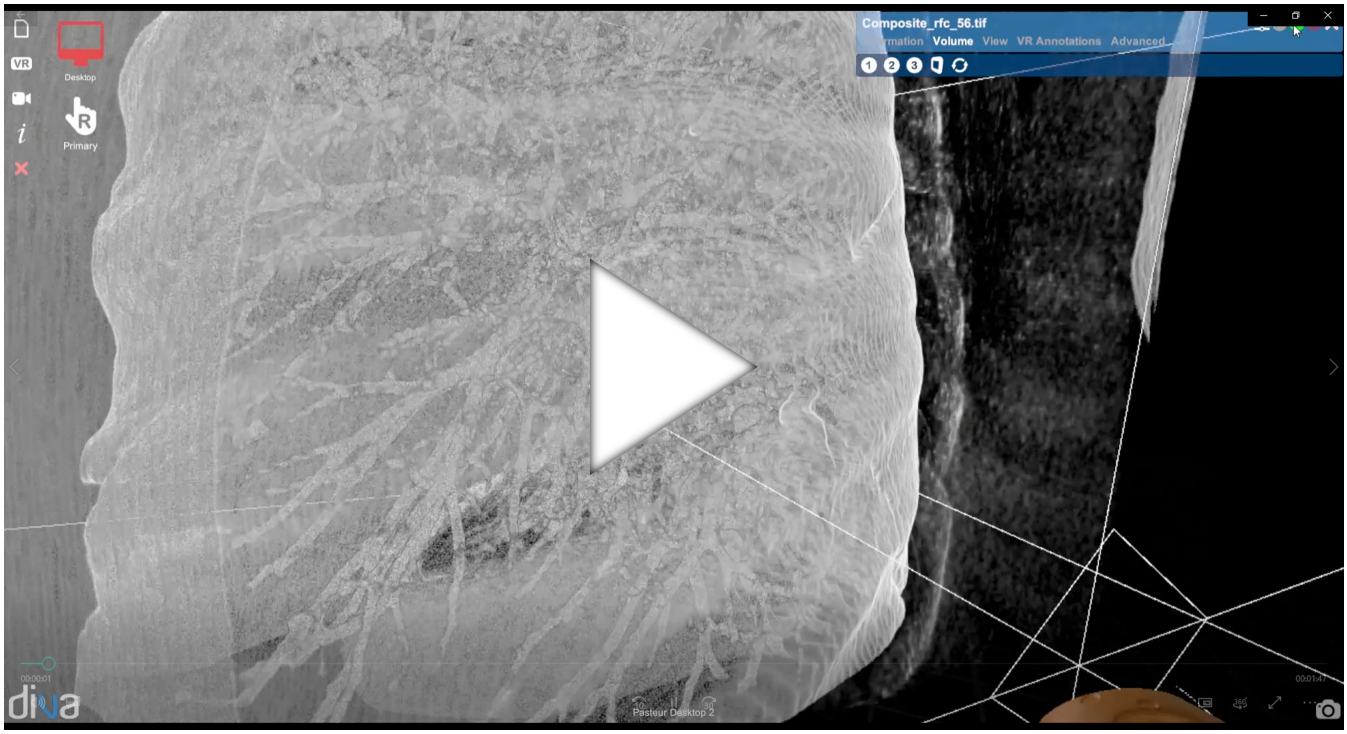
7 VIDEOS



Video S1. Annotation in DIVA on confocal microscopy images of mouse olfactory bulb interneurons. The whole pipeline is comprised of successive steps. **1)** Tagging voxels in virtual reality. Overlay of raw data and tagging data with positive and negative tags respectively in cyan and magenta. **2)** Model training (here RFC) using DIVA adapted interface. **3)** Inference on the whole stack. Overlay of raw data and output probabilities for RFC. Transfer function is finally set for appropriate visualization.



Video S2. Tagging step in DIVA on CT-scan of lung tumor from the Medical Segmentation Decathlon (MSD) challenge : lung_001. The transfer function is first set to ensure proper visualization of the tumor in desktop mode. The tagging is performed in VR using the clipping plane tool to navigate inside the volume and grasp the contour of the structure of interest. Overlay of raw data in gray and tagging data with positive and negative tags respectively in cyan and magenta.



Video S3. Visualization of resulting annotation in DIVA on CT-scan of lung tumor (white arrow) from the Medical Segmentation Decathlon (MSD) challenge : lung_003. The file here has three channels, the first for the CT-scan in gray-scale, the second for the experts' segmentation in green, and the third for the inferred annotation in blue to red color-scale. These three channels are alternatively turned on and off to ensure proper visualization in VR of the structure of interest and its proposed segmentation. After using the desktop clipper, a zoom is performed on the tumor.

frontiers in Bioinformatics

METHODS
published: 20 January 2022
doi: 10.3389/fbinf.2021.775379

Towards Human in the Loop Analysis of Complex Point Clouds: Advanced Visualizations, Quantifications, and Communication Features in Virtual Reality

Thomas Blanc[1,2], *Hippolyte Verdier*[3], *Louise Regnier*[1,2], *Guillaume Planchon*[3], *Corentin Guérinot*[3,4], *Mohamed El Beheiry*[3], *Jean-Baptiste Masson*[3*†] and *Bassam Hajj*[1,2*†]

[1]Laboratoire Physico-Chimie, Institut Curie, PSL Research University, CNRS UMR168, Paris, France, [2]Sorbonne Universités, UPMC Univ Paris 06, Paris, France, [3]Decision and Bayesian Computation, CNRS USR 3756, Department of Computational Biology and Neuroscience, CNRS UMR 3571, Université de Paris, Institut Pasteur, Université de Paris, Paris, France, [4]Sorbonne Universités, Collège Doctoral, Paris, France

OPEN ACCESS

Edited by:
Shamit Soneji,
Lund University, Sweden

Reviewed by:
Daniel Haehn,
University of Massachusetts Boston, United States
Jan Byška,
Masaryk University, Czechia

***Correspondence:**
Jean-Baptiste Masson
jean-baptiste.masson@pasteur.fr
Bassam Hajj
bassam.hajj@curie.fr

[†]These authors have contributed equally to this work and share last authorship

Specialty section:
This article was submitted to Data Visualization, a section of the journal Frontiers in Bioinformatics

Received: 13 September 2021
Accepted: 24 December 2021
Published: 20 January 2022

Citation:
Blanc T, Verdier H, Regnier L, Planchon G, Guérinot C, El Beheiry M, Masson J-B and Hajj B (2022) Towards Human in the Loop Analysis of Complex Point Clouds: Advanced Visualizations, Quantifications, and Communication Features in Virtual Reality. Front. Bioinform. 1:775379. doi: 10.3389/fbinf.2021.775379

Multiple fields in biological and medical research produce large amounts of point cloud data with high dimensionality and complexity. In addition, a large set of experiments generate point clouds, including segmented medical data or single-molecule localization microscopy. In the latter, individual molecules are observed within their natural cellular environment. Analyzing this type of experimental data is a complex task and presents unique challenges, where providing extra physical dimensions for visualization and analysis could be beneficial. Furthermore, whether highly noisy data comes from single-molecule recordings or segmented medical data, the necessity to guide analysis with user intervention creates both an ergonomic challenge to facilitate this interaction and a computational challenge to provide fluid interactions as information is being processed. Several applications, including our software DIVA for image stack and our platform Genuage for point clouds, have leveraged Virtual Reality (VR) to visualize and interact with data in 3D. While the visualization aspects can be made compatible with different types of data, quantifications, on the other hand, are far from being standard. In addition, complex analysis can require significant computational resources, making the real-time VR experience uncomfortable. Moreover, visualization software is mainly designed to represent a set of data points but lacks flexibility in manipulating and analyzing the data. This paper introduces new libraries to enhance the interaction and human-in-the-loop analysis of point cloud data in virtual reality and integrate them into the open-source platform Genuage. We first detail a new toolbox of communication tools that enhance user experience and improve flexibility. Then, we introduce a mapping toolbox allowing the representation of physical properties in space overlaid on a 3D mesh while maintaining a point cloud dedicated shader. We introduce later a new and programmable video capture tool in VR and desktop modes for intuitive data dissemination. Finally, we highlight the protocols that allow simultaneous analysis and fluid manipulation of data with a high refresh

rate. We illustrate this principle by performing real-time inference of random walk properties of recorded trajectories with a pre-trained Graph Neural Network running in Python.

Keywords: virtual reality, point clouds, inference, microscopy, human in the loop, 3D maps

INTRODUCTION

Scientific research is producing large amounts of data of various types and dimensionality. Exploring this increasingly complex data is essential to guide experiments, rapidly extract relevant information, and guide future developments. Multidimensional point cloud data are generated in several scientific fields, including LIDAR, computer-assisted design, segmented medical images, and electron microscopy, to name a few. This paper shows an application centered on point clouds generated from single-molecule localization microscopy (SMLM) (Betzig et al., 2006; Rust et al., 2006) without losing the generality of the proposed method solutions.

In SMLM, single-molecules are observed in their natural cellular environment to extract information related to their dynamics and interactions in live cells (Höfling and Franosch, 2013; Manzo and Garcia-Parajo, 2015), or to reconstruct super-resolution images of cell organelles (Betzig et al., 2006; Rust et al., 2006). Large-scale recordings of single molecules at the nanometer resolution reveal the complex interplay between 4D geometry (space and time) and biological activity. Understanding these complex time-evolving structures requires freely navigating within the data and analyzing portions of it.

Another significant component associated with single-molecule imaging is the complexity of the data. Positions and dynamics of biomolecules are accessed through multiple preliminary operations (Chenouard et al., 2014), such as image deconvolution and tracking, which leads to noisy datasets. Furthermore, these datasets are usually highly heterogeneous in space, time, and other associated properties. Furthermore, organelles unseen in datasets influence the geometry of the recorded point cloud in the surrounding environment. Hence, we are convinced that numerous scientific initiatives benefit from immersive visualization modalities such as virtual reality (VR) to enhance data comprehension. Data in this context may be uploaded to a VR environment and visualized using a custom shader program linked to relevant experimental parameters.

Exploring the data in VR offers the possibility to grasp the global geometrical structures while allowing the possibility to dive within the data to explore more local structures and their relation in space. Besides, the interaction in VR facilitates the quantitative analysis of the data through the combination of user intervention and optimized analysis algorithms. The ability to easily and quickly isolate regions of interests in 3D for example is time consuming and impractical even with efficient desktop software such as ViSP (Beheiry and Dahan, 2013), while these sorts of interactions are faster and more precise using VR. Several applications including our software platform Genuage are dedicated to visualize and analyze point cloud data in virtual reality (Blanc et al., 2020). Genuage is an open-source VR–compatible platform for visualization of n-dimensional point cloud data. It features a set of VR-assisted tools for data exploration, data selection, and built-in analysis. While VR visualization parameters can be adapted to the various types and dimensionality of the recorded point clouds, quantification, on the other hand, seems far from being standard. Point clouds generated in different acquisition modalities can vary in dimensionality and might underlay specific information that requires dedicated algorithms for analysis. Moreover, such algorithms are developed in different scripting platforms and remain case-dependent. The standardization of the quantification process remains thus difficult. In addition, complex analysis can require large computational resources, making the real-time VR experience uncomfortable. Moreover, visualization software is mostly designed to represent a set of data points but lacks flexibility in manipulating and analyzing the data. Examples include generating new data properties to visualize, calculating physical properties from selected points, creating new point clouds derived from the one already uploaded and combining data sets.

This paper presents the Genuage software for point cloud visualization and interaction and highlights recent developments. We introduce new approaches and processes implemented within the Genuage platform, in which enhanced flexibility and user experience are made possible by a set of communication tools and built-in functions. We discuss communication protocols with user-defined libraries applied on a selection or an entire data set. This communication mode allows DLLs to compute specific properties on selected data, generate new data sets or create a new point cloud. We demonstrate this approach by analyzing the blinking of single-molecule localization data where it is desirable to combine localizations originated from blinking molecules, to compute local density, or to determine dynamical properties. We introduce a new set of tools for calculating and presenting physical properties in 3D overlaid to a 3D mesh while maintaining a point cloud dedicated shader. We additionally introduce learning procedures on graphs and their relation to points clouds in a biological context. We show the interaction with data performed while maintaining a high refresh rate within the VR headsets, thus maintaining user comfort when communicating with other scripting platforms even if large computational resources are required. We illustrate this principle by performing real-time inference of random walk properties of recorded trajectories with a pre-trained Neural Network running in Python. Finally, we introduce a new programmable video capture tool in VR and desktop modes for intuitive data dissemination.

METHODS

VR for Point Cloud Visualization

VR is starting to play a more important role in research initiatives. From cardboard inserts intended for smartphones to advanced headsets with multiple sensors for positional tracking, VR contributes to the design of projects for various science topics.

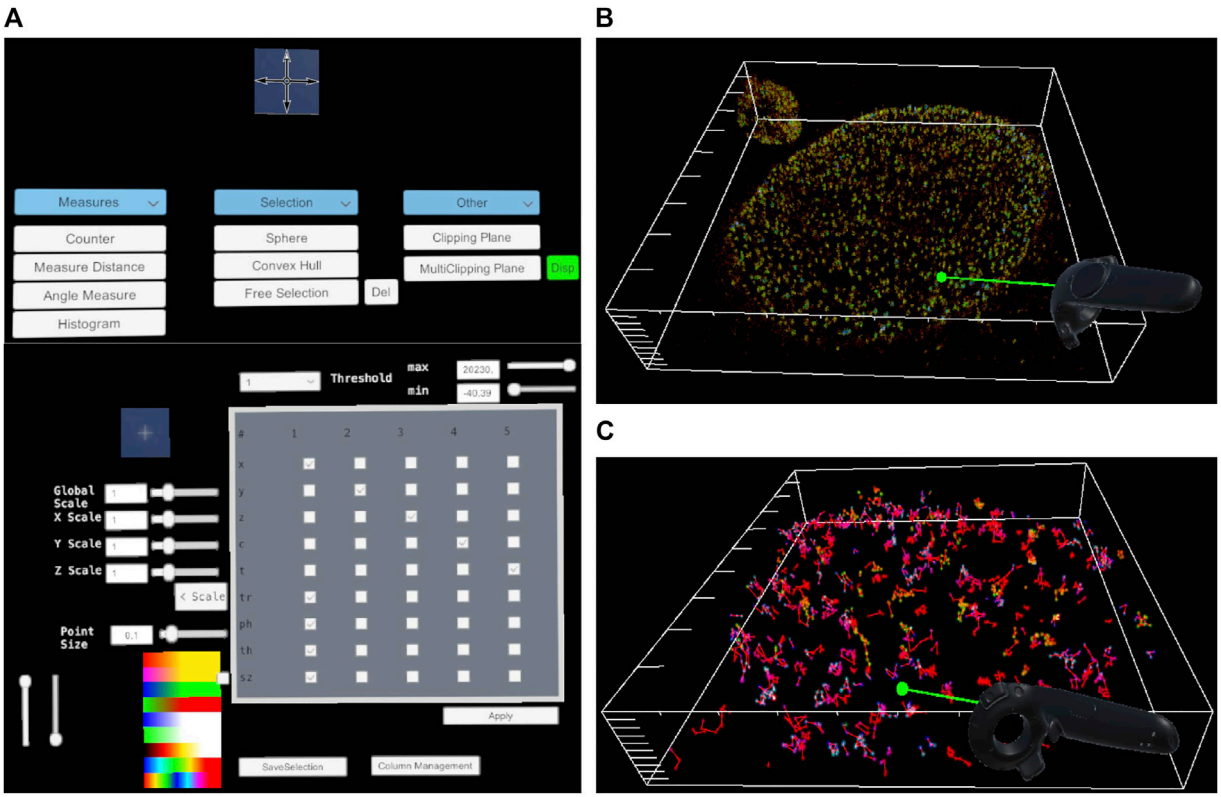


FIGURE 1 | General overview of Genuage. The VR interface **(A)** allows the adjustment of the visual parameters associated with the point cloud. It includes associating visual parameters to specific columns in the data, adjusting the color code, scaling, and aspect ratio. A set of analysis tools allow performing VR-assisted measurements and complicated 3D selections. Clipping plane tools allow us to explore the inner layers of the point cloud in VR at arbitrary angles. Multidimensional point cloud data generated by single-molecule localization microscopy are explored in VR in 3D **(B)** and (3D + time) **(C)**.

Facilitating this process are freely accessible development platforms, such as Unity and Unreal Engine, and the availability of many open-source projects. VR applications can now be a part of the teaching (Brůža et al., 2021) and research environments (Matthews, 2018; El Beheiry et al., 2019).

Applications centered on point clouds span various fields. Some relate to data tagging with machine learning such as those in (Bergé et al., 2016; Stets et al., 2017; Wirth et al., 2019; Liu et al., 2020; Ramirez et al., 2020). They range from optimized interfacing to easy interactions with the data to mixed interactions with various algorithms to prepare data or detect specific structures of interest. Some hybrid applications can be found in astronomy such as (Gaia Sky, 2014). General visualization and interaction software include PointCloud XR and developments centered on compression to ensure visualization in VR (and AR) (Pavez et al., 2018). Other applications such as MegaMol and UnityMol are dedicated for the visualization of point-based molecular data sets (Doutreligne et al., 2014; Grottel et al., 2015). Within the context of microscopy and biological data analysis, which is the focus of this paper, the open source vLUME software (Spark et al., 2020) provides a direct way to represent, interact, and analyze static point clouds from single-molecule experimental data. It features numerous interfaces and high-quality visualization to highlight structures and better understand biomolecule geometric distributions within the cell. We provide in the supplementary a detailed feature comparison table.

Genuage for Point Cloud Representation and Analysis

General Features of Genuage

We recently introduced the open-source software platform Genuage (Blanc et al., 2020; Genuage GitHub, 2020) to address visualization, interaction and analysis of complex multi-dimensional point cloud data. The platform can accommodate several types of point cloud data. Genuage is built on the Unity game engine and was initiated on the simple premise that extracting information from high dimensional point clouds is a non-trivial task. VR immersion provides the dual possibility of increasing the number of dimensions represented and allowing more direct interaction with data. Genuage has a dual interface: a computer desktop interface and a VR interface. The desktop interface derives its design principle from ViSP (Beheiry and Dahan, 2013): it allows users to load the data and set general representation parameters (**Supplementary Figure S2**). While all these parameters could be defined within the VR interface, data visualization and analysis practices in the lab are often incompatible with long times passed

within the VR environment. Hence, we developed a dynamic process where the user can switch from one interface to the other based on the specificity and complexity of the action, they wish to execute following the same design principles of DIVA (El Beheiry et al., 2020). The VR interface allows visualization and navigation within the point cloud (**Figure 1**; **Supplementary Figure S3**). It provides multiple means of interaction and for adjusting visualization parameters within the VR environment. In addition, various tools for measuring distances, angles, counting, and time series analysis are accessible.

The data sets and associated parameters presented in this paper stems from single-molecule localization microscopy experiments and present some specificities that we will highlight in the following paragraphs.

Single-Molecule Experiments and Data

Fluorescence microscopy is a powerful tool in modern biological studies. It provides a window to observe the inner functioning of cell organelles in physiological conditions with high specificity. In conventional fluorescence microscopy, images of the whole sample are acquired with a limited resolution due to the optical diffraction limit (Lichtman and Conchello, 2005). However, imaging individual molecules allow overcoming this limit. Individual labeled proteins can be imaged and localized with high precision in live and fixed cells thanks to bright fluorescing molecules. Imaging sparse distributions of single molecules over time and localizing their centers with high precision provides two types of information. First, it offers a means to follow the dynamics of individual emitters in live cells. The quantification of each molecule's diffusion behavior and interaction kinetics provides a statistical map of molecular interactions avoiding thus population averaging. Second, it allows the reconstruction of super-resolution images with a resolution below the diffraction limit of the microscope. The principle relies on splitting the emission of molecules in time and space such that only a sparse subset of molecules is active in a single acquisition frame. Each molecule is localized with a precision way below the diffraction limit. Repeated steps of localization, bleaching, and activation of a new subset of single-molecules allow to sample the whole structure of interest and reconstruct a pointillist image in super-resolution. In both cases, the output of the localization algorithm is a set of 3D coordinates [e.g., (Hajj et al., 2014; Nehme et al., 2020)], and, depending on the complexity of the detection microscope, several associated parameters such as the color, the intensity of the detected molecule, the time stamp, and the molecular orientation to name a few. Visualizing and analyzing such complex datasets is not trivial on a 2D screen given the large dimensionalities of the recorded point cloud. The complexity and the density of data points can conceal or intersect with other features within the data. In addition to the visualization aspects, analysis might require specific tasks to be performed within the complex dataset such as selection and counting. Users often face the need to analyze a specific trajectory or a selection of trajectories within a region of interest. Thus, the main requirements for exploiting single-molecule localization data are an efficient and intuitive perception of multidimensional data, ease of navigation through the data, and a facilitated interaction to accomplish specific tasks such as placing landmarks for measurements or counting and performing selections.

There is a clear advantage in using Genuage and VR to navigate and interact with such complex data. However, analyses of single-molecule datasets are far from being standard. It is common for different users to adopt different strategies for the quantification of the recorded cloud. Custom-made algorithms are often required and can continuously improve with evolving functionalities. Moreover, different algorithms might impact the original point cloud differently, for instance by generating additional properties for the point cloud or by creating a completely new data set. There is thus a clear need to adapt Genuage for different custom-made algorithms with non-standardized outputs.

General Framework for Visualization and Analysis in the Genuage Platform

A table of 3D coordinates defines the point clouds and other associated parameters generated by a specific experimental and analysis approach. Genuage can read point cloud data from any text file with any separator type providing that the data is organized in rows and disregarding the presence of any header. Individual points are represented directly with geometric shaders as Gaussian shapes with adjustable standard deviations to provide an adequate "size" to the points within the VR environment. The geometry shader spawns 4 points per data point to define a sprite with the Gaussian shape. Characteristics of Gaussian profiles such as color, intensity, or transparency can be mapped to specific data columns associated with a set of point cloud parameters or obtained as a result of performed analysis. Different color code palettes, including colorblind friendly palettes, are provided to account for different experimental data requirement and avoid visualization artifact (Borland and Taylor Ii, 2007). Subsets of points are visible or invisible based on specific thresholding parameters with nearly no impact on software performance or framerate. In addition, other high-dimensional information can be presented, such as vectorial data associated with points. For example, molecular orientations appear as an overlay on the point cloud of individual orientation-dependent color-coded segments. Similarly, we can display dynamic (time series) point clouds as a function of the recorded time or scrolling time windows. Trajectories are linked via segments in 3D, as can be seen in (**Figure 1C**). The user fully controls the mapping and correspondence between data columns and visual parameters. In (**Figure 1A**) we show a few components of the interface that are directly related to the discussed application.

Visualizing dense 3D point clouds can be challenging as the user's relative position concerning the observed reconstructed data may not fully visualize the underlying structure. Similarly, highly anisotropic data can be challenging to visualize due to orientation dependency. In Genuage, the user can render a set of point cloud data visible or invisible based on thresholding specific parameters present in the data. Also, thanks to a clipping plane feature, the user can control within the VR environment a 3D

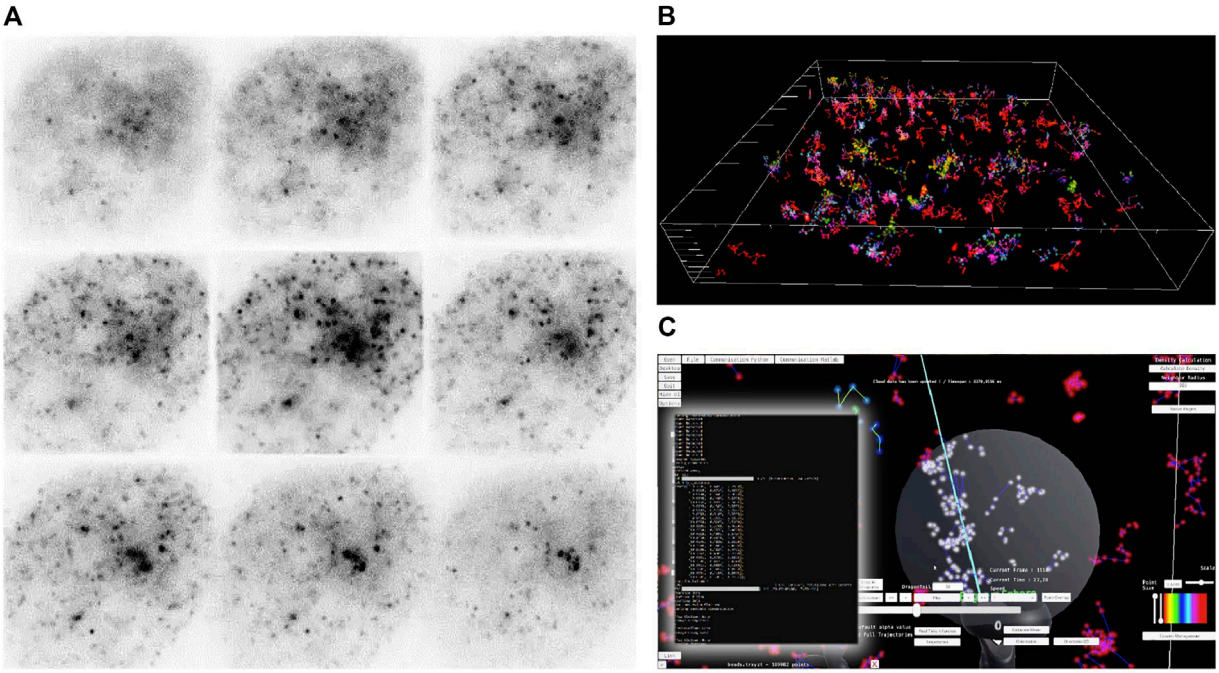


FIGURE 2 | Workflow for single-molecule analysis in Genuage. **(A)** Raw images of diffusing particles in the nucleus of a cell are captured with MFM. Images are analyzed, and a table is generated containing the position of each particle and the corresponding timestamp. **(B)** The localizations are imported into Genuage and explored in VR. Color codes can be tuned to represent individual trajectories or the temporal evolution of the point cloud dynamics. **(C)** A screenshot of the Genuage interface as a selection of point cloud subsets and analysis is performed. Trajectory selection is performed using the VR controllers with a visual aid in the form of a sphere. Other shapes can be used to select points. In highly convoluted point distributions, the user can define landmarks generating a convex hull that will capture points of interest. Results are displayed as changes of the trajectory colors, with intensity depending on the value of the sub-diffusive exponent. The inference analysis is performed in real-time on the selected trajectories (inset shows real-time calculation performed). Various means to display results are available. The screenshot also shows the interface in the current version of the Genuage software. Most of the buttons and displays are self-explanatory [an introduction to the design can be found in the videos associated with (Blanc et al., 2020)].

plane that conceals the points it intersects. Furthermore, this clipping plane can be held in a place and combined with several additional ones to allow exploration of dense point cloud data at arbitrary angles.

In addition to the visualization aspects, we provide several built-in quantification tools in Genuage. VR is a vehicle to perform otherwise time-consuming quantification tasks more precisely. A set of essential tools include 3D length measurement, counting, angle measurement, profiling the point cloud along a segment as a bin count, selecting a subset of the point cloud, and exporting data. All geometric operations in 3D space are simplified thanks to the VR environment. JSON files are generated as a means to save the manipulation progress and to record the analysis results.

Advanced analysis such as diffusion coefficient inference on selected time series data complements the essential tools that are presented. In addition, a new tool to overlay different shaders is implemented to represent local physical properties computed on the local point cloud data set. For example, we provide an analysis tool for creating a complex 3D mesh based on local point cloud density. In each region, the code infers the local diffusion behavior of dynamic molecules data and presents it as a variation in the color of the mesh or as surfaces with adequate color. The implementation will be further detailed in the following section.

The Genuage platform is developed as a versatile package for point cloud data and is highly modular to allow easy repurposing of portions of the code. Besides visualization, the goal of Genuage is efficient, and simplified data analysis. As point cloud data is ubiquitous, developing a general analysis framework is challenging. Our first development ensured compatibility with Python with a dedicated communication protocol launchable from Genuage. This development was motivated by ensuring that complex analysis could be performed outside the platform, especially if they required significant computational resources. We used a similar approach for volumetric data annotation in the DIVA-cloud platform (Voxel Learning in DIVA, 2020). Efficient uses of the pipeline consist of transmitting a subset of selected data to a python script for analysis and using the results to enhance the visualization of this subset or provide extra information in space and time. We also implemented the Matlab pipeline as it is a popular tool for biological data analysis. Similar pipelines will soon be added for R and Julia. An example of an application is shown using a custom Graph neural network to perform single-molecule trajectory analysis (see **Figure 2**; **Supplementary Video S1**).

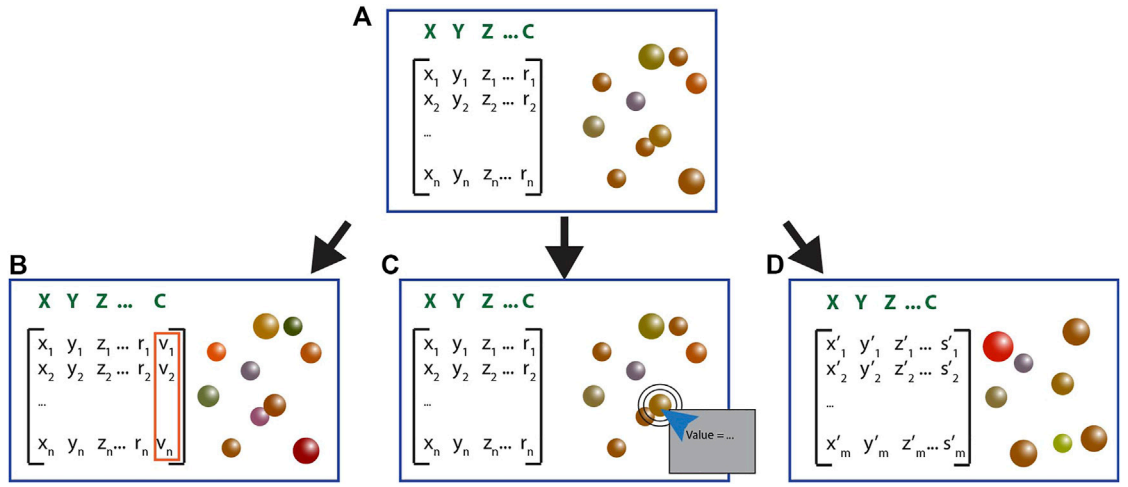


FIGURE 3 | Different library configurations are compatible with Genuage and allow performing specific manipulation of the data. **(A)** Initially, the point cloud is imported as a table of coordinates displays in Genuage as 3D spheres positioned at (X,Y,Z) coordinates with specific visual attributes like colors (c). Depending on its format, the DLL: **(B)** performs a specific calculation on the point cloud that results in an additional data column that could be assigned to a specific visual feature, **(C)** retrieves the information from selected points. It displays the resulting calculus in a popping up window, **(D)** generates an entirely new point cloud with different elements based on specific mathematic manipulation.

Our newest development focused on numerical data analysis that either requires limited computational resources or requires a complex set of packages. We added a toolbox to import ad-hoc external DLLs libraries. While we use the classical development process of using addons, plugins, and DLLs to enhance the Genuage platform, our challenge lied in the end-user community willingness to engage in the challenges of developing VR software and in their limited computing facilities (usually computers with basic graphic cards and few CPU cores). We streamlined the development and use of a VR software for experimental scientist and biologists. Furthermore, the architecture is optimized to ensure usability on most computing platforms. We provide different templates for the required possible formats of the DLLs and structures. Different DLL formats allow performing specific tasks. Currently, we provide 3 configurations of libraries that allow efficient integration within the Genuage platform (**Figure 3**; **Supplementary Video S2**). These configurations provide guidelines for a user to develop other possible integration architectures. We provide code templates to help user easily create DLLs from their analysis algorithms, and integrate their code into an analysis routine through Genuage. Users are able to run custom algorithms with no need for advanced knowledge in C# or software development. We favored three analysis configurations. First, the library generates new columns in the point cloud data represented within the VR environment and associated to a specific visual parameter (**Figure 3B**). Second, the library generates numerical outputs shown in an extra visualization window (**Figure 3C**). Third, the library creates a new point-cloud dataset (**Figure 3D**). To illustrate these procedures and provide development guidelines for new libraries, we provide three DLLs. The first one simplifies point density estimation. The second one computes the apparent diffusion coefficient on a selected trajectory. The third one combines and displays point clouds localizations generated by the same molecule due to the inherent molecular blinking commonly encountered in single-molecule super-resolution acquisitions (check **Supplementary Information** for more details).

While VR is instrumental for data analysis, results are often communicated without the VR environment. Furthermore, VR and AR remain new technologies that are not widespread. Hence, we developed a toolbox to capture information within the VR environment and export it in an easy to share format. We developed a dedicated toolbox that features an image capture module within the VR environment and a video capture mode where the user defines landmarks for the displacement of a virtual camera within the virtual point cloud. Furthermore, the user can adjust the timing at each waypoint and set visual parameters such as the color and thresholding. Then, the user can access in desktop mode the newly created external file. The file can be further adjusted manually to define the position of the waypoints and the animation using a dedicated scripting language. In addition, the desktop mode has an interface for timing adjustment. As a showcase, we present the video capture of dual color super-resolution localization microscopy data sets performed in budding yeast (**Supplementary Video S3**).

Advanced Analysis

Point clouds are ubiquitous and thus analysis can vary significantly in complexity (Charles et al., 2017; Qi et al., 2017; Liu et al., 2019). These last 5 years, analysis of point clouds have been the center of numerous efforts in machine learning. The first challenge of any set of point clouds analysis is the data's variable size, preventing usual machine learning approaches that require data standardization. The second is the complexity of possible

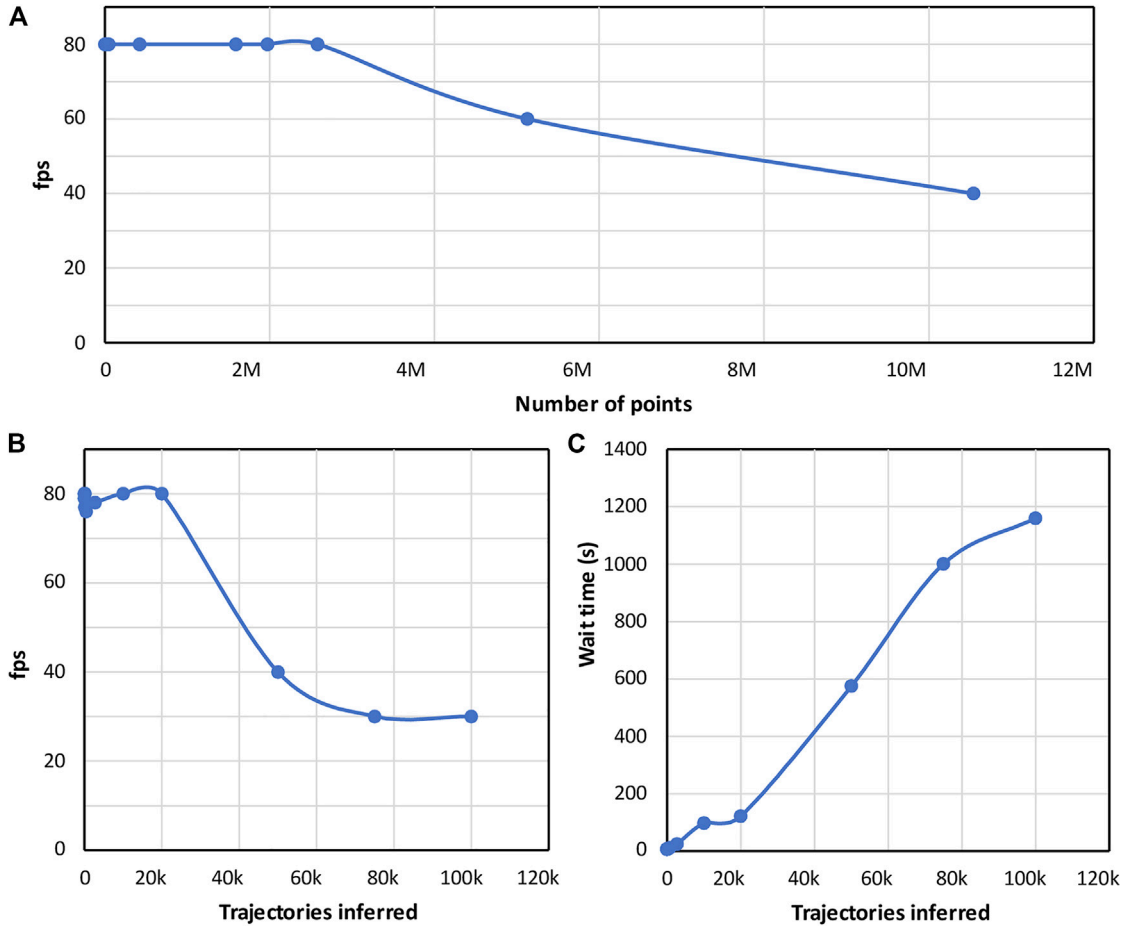


FIGURE 4 | Performance quantification: **(A)** The frame rate of the graphic card is reported as a function of the number of points that are visualized in VR. **(B)** Evolution of the frame rate of the graphic card as a function of the number of the trajectories that are analyzed using the pre-trained neural network. A frame rate of 80 Hz as imposed by the used VR headset defines the range for a comfortable utilization of the VR headset. The refresh rate remains at a comfortable level for up to 30000 trajectories. Such number is considered extremely high compared to realistic data sets. **(C)** Evolution of computation time of the pretrained neural network as a function of the number of inferred trajectories. The reported analysis time is extremely fast regarding the computation required by the GNN. The performance quantifications are performed for a computer with the following specs: Intel Core I5-8400 CPU, 16 Gb ram, Geforce 1080Ti GPU.

underlying models (if they exist) and associated challenges in performing Bayesian analysis (Jaqaman et al., 2008; Höfling and Franosch, 2013). Our domain of application in this paper matches these two challenges. We cannot control time series size, and most underlying models have either no known likelihood formula or are computationally intractable. We refer the reader to (Höfling and Franosch, 2013; Manzo and Garcia-Parajo, 2015; Miller et al., 2017) for a general introduction. Hence, we used simulation-based inference to analyze selected data within the virtual environment. We leverage our recently developed graph neural networks (Fey and Lenssen, 2019; Verdier et al., 2021) to run efficient parameter inferences. We used the following architecture (**Supplementary Figure S1**, with explication in the caption). The inference model was trained on numerical datasets of various random walk using the "ANDI" package (AnDi, 2020; Muñoz-Gil et al., 2020). The model is pre-trained, and the number of associated parameters is reasonable for a deep model (300,000–500,000). VR representation was done on a GPU card while the inference was run directly on the CPU, leveraging multi-threading to analyze sets of individual trajectories as they were selected by the user.

In the context of single-molecule tracking experiments, a common task is to find the nature, and associated parameters of dynamical random walks. The underlying idea is that the random walk is a proxy to the physical interactions between the biomolecules and their cellular environments. For this application, typical analysis time is counted in multiple hours. From the localization process of single molecules in images to analysis of trajectories, it can take from hours to multiple days (especially on a single computer). Here, the computation time of the trajectory analysis is massively reduced using a pretrained graph neural network. In our implementation, the challenge was to run an analysis with the pretrained GNN while the representation in VR was running and to analyze the selection of point clouds that could belong to a single or a set of trajectories. Furthermore, the analysis should not significantly slow down the

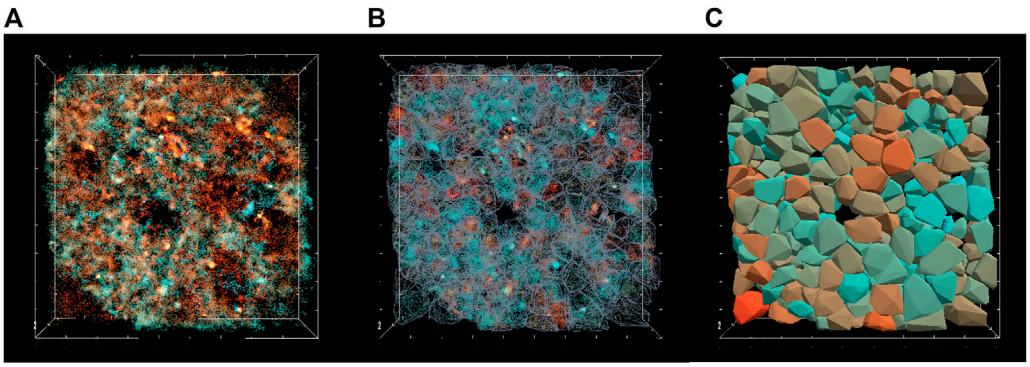


FIGURE 5 | The principle of 3D diffusion map analyses. **(A)** Experimental point cloud data generated from beads injected, imaged in 3D and tracked in the nucleus of a living cell. The panel shows all the recorded localizations color coded by the axial position. **(B)** A 3D mesh allows identifying different regions with a sufficient number of point clouds to perform the analysis. **(C)** in each region, the diffusion coefficient is inferred and shown as a closed surface color coded by the diffusion coefficient.

refresh rate to avoid discomfort when analyzing and exploring the data. We found that the rendering performance in Genuage was not impacted when performing the analysis with the pretrained GNN applied on a selection of trajectories. The refresh rate remained at a comfortable rate (Zhang, 2020) of 80 Hz as recommended by the testing VR headset. To test the limits, we quantified the refresh rate as a function of an increasing number of selected trajectories. We found that the refresh rate remained at a comfortable level for selections containing up to 30000 trajectories. Such number is considered extremely high compared to realistic data sets. While the GNN analysis could take up to few minutes to be performed on such large number of trajectories, the VR experience is not impacted, and the user is able to interact with the data during this waiting time. The quantification of the performance is reported in **Figure 4**.

Nevertheless, when possible, Bayesian Inference remains a powerful analysis tool, especially for mapping spatially physical properties of the cell. One possible analysis approach is determining a landscape of diffusivity and interaction potential based on the recorded point cloud trajectories, including the very short ones, within a given area (Masson et al., 2014; Beheiry et al., 2015) (check supplementary for details). Here we extended this principle to 3D for point cloud data generated thanks to the volumetric imaging capability of MFM. We implemented a 3D mapping of effective diffusion and forces in 3D through a DLL as a proof of concept. First, clusters of point clouds are identified using a ***k***-means clustering algorithm (MacQueen, 1967), and a 3D mesh is generated to define the borders of the different regions. The assumption is that physical properties, such as diffusion coefficient and potential, remain constant in small regions. Second, a Bayesian inference approach estimates the local diffusion coefficient and other physical properties (check supplementary). Finally, the calculated values are uploaded as additional columns. From a visualization point of view, the new challenge here resides in overlaying the generated regions and the associated properties to the point cloud in VR. We opted for a wireframe visualization approach for discriminating the generated 3D mesh and a 3D surface presentation of the associated physical property (**Supplementary Video S4**). The color code and transparency of each region are adjusted to reflect the calculated physical properties. **Figure 5** illustrates the analysis and visualization pipeline.

RESULTS: PROOFS OF CONCEPT AND APPLICATIONS

Here, we demonstrated our procedures on a set of point cloud data consisting of single-molecule localizations and trajectories of beads injected in the nucleus of living cells. U2OS cells were plated on microscope coverslips. Particles 50 nm in size were injected into the nucleus and allowed to freely diffuse. Volumetric images were recorded at 30 ms exposure time using multifocus microscopy (MFM) (Hajj et al., 2014). Thanks to the addition of several diffractive optical elements on the emission path of a widefield microscope, MFM captures instantaneous volumetric images of the whole cell nucleus by tiling different focal planes side by side on the same camera. MFM fluorescence images are reconstructed as Z-stacks in a post-processing step. The analysis pipeline retrieves the 3D sub-pixel localization of each particle *via* 3D Gaussian fitting. Trajectories are later generated using the Utrack algorithm (Jaqaman et al., 2008) (**Figure 2B** and **Figure 6A**). Upon injection, the particles explore the heterogeneous volume of the nucleus, and the diffusive behavior is thought to be a direct indicator of the physical properties and compaction of the surrounding chromatin. In addition, and regardless of the dynamical aspects, the accumulation of all the localizations of the single particles reconstitutes a super-resolution image of the empty space that is surrounded by the chromatin fibers.

In **Figure 2** we show a capture of the analysis with the Genuage interface. **Supplementary Video S1** shows a full demonstration of the inference being performed sequentially on subsets of selected trajectories. They also show the entire procedure from loading the data to performing a complete analysis of experimental data and procedures to import custom analysis software within Genuage. Analysis time is dependent on the size of the selected data. As a graph is associated with all the selected trajectories, the

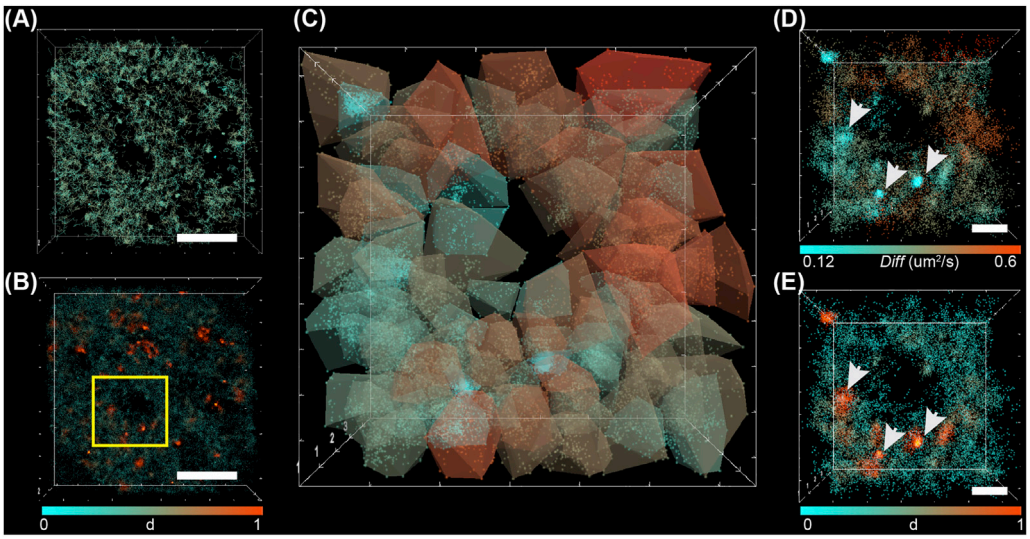


FIGURE 6 | Pipeline for calculating 3D diffusion map of point cloud dynamic data. **(A)** Reconstructed 3D trajectories of beads injected in the nucleus. **(B)** A slice in the reconstructed volume shows a point cloud color-coded by the local density (***d***). A very heterogeneous density is observed pointing towards different chromatin compaction. **(C)** 3D diffusion map generated over a small nucleus area represented by the blue square in **(B)**. The different areas are presented by surfaces with controllable transparency and color-coded by diffusion (*Diff*). **(D)** To easily visualize the diffusion landscape point clouds are colored by the diffusion coefficient. The arrows point toward examples of low diffusivity area where particles are assumed to be trapped. This hypothesis is supported by the corresponding high local density estimation as presented by **(E)**. Scale bar: 5 microns in **(A)** and **(B)**, 1 micron in **(D)**.

distribution of trajectory lengths translates into time distributions in receiving the analysis results.

In the **Supplementary Figure S6**, we show the measured evolution of the framerate with increasing analysis constraints. Beyond 30,000 points, we see a significant decrease in the VR framerate and an increase in the waiting time to receive the analysis result.

Thanks to Genuage and the new set of communication tools, we generated a density map of the particles within the nucleus using an external DLL (provided). It is a direct link to local constraints imposed by the surrounding DNA. **Figure 6B** shows a slice through the recorded volume where points are color coded by local density. Observed empty spaces are a direct hint for the presence of a chromatin area with high compaction or nuclear compartment that large particles cannot bridge. Conversely, high-density regions point towards trapped particles due to local chromatin density.

To check how compaction affects dynamics and the compaction hypotheses directly linked to diffusion, we analyzed the dynamics of single particles. Real-time analysis allowed us to efficiently explore the complex nuclear environment to retrieve local diffusive properties. As a result, we have observed different diffusion and sub-diffusion exponent parameters in various areas of the nucleus, providing evidence of different compaction levels.

To expand further our investigation, we generated diffusivity maps all over the observed volume and correlated it to the presence of a restricted access area or high particle density area. **Figures 6C,D** shows a 3D diffusion map generated for a defined area of the nucleus. A very heterogeneous diffusion landscape was observed with intermingled regions of slowly diffusing particles and trapped ones. It is linked to local chromatin compaction and was confirmed by calculating the local density of point cloud **Figure 6E**.

Future works will focus on disturbing the chromatin organization locally or on a large scale and monitoring the evolution of the diffusion. We think that the VR experience and advanced real-time analysis within Genuage will provide a valuable tool for guiding our future findings.

DISCUSSION

Virtual and augmented reality will likely play a more critical role in research and medical application. Large scale initiatives using these technologies will fuel the acceptance of the technologies and improve their usability. While initiatives relying on VR to improve complex data visualization are more familiar with applications on microscopy-based data or medical images, we feel that leveraging VR as a part of the analysis pipeline will be a crucial element pushing for research adoption of the tech.

This paper shows dedicated software for real-time visualization and analysis of multidimensional point cloud data such as those generated by single-molecule microscopy. We show that VR is an efficient tool to visualize and interact with complex point-cloud data. Visual features can represent high dimensional data sets and the corresponding output following a quantification process. Genuage is an optimized tool to facilitate data analysis of complex non-isotropic point clouds such as those obtained from biomolecule dynamics.

We have shown a set of libraries and toolboxes implemented in the Genuage platform to perform human-in-the-loop point

cloud analysis in VR. Our developments focused on allowing analysis without impacting the image rendering refresh rate and thus ensuring fluid interactions between the user and the data. We provided the required tools to allow efficient data extraction from the VR environment to be displayed onto regular 2D viewing devices. We ensured compatibility with popular analysis frameworks while keeping on adding new ones. Finally, we developed proof of concept analysis pipelines to demonstrate the analysis process in VR and provide guidelines to develop new ones. We chose an application deeply linked to biomolecule research. Nevertheless, the generality of the approach allows applications to any kind of point cloud data.

Our future initiative will center on four topics. First, we will develop Genuage-cloud, a platform extension to allow complex analysis to be run on the cloud through a Django interface and entirely run the visualization and analysis in the cloud while streaming the images to the VR headset. Second, we will extend the Genuage platform to run Augmented Reality (AR) applications on glasses and tablets with a new library focused on amortizing the representation of large point clouds on devices with limited memory. In addition, we will develop new shaders to render the fusion of different data types in the VR environment, facilitating data interpretation during the analysis. Finally, we will develop a physical engine library to allow mixing user interaction and physical simulations to explore organelle structure dynamics in the cell.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/Genuage/Genuage.

AUTHOR CONTRIBUTIONS

TB, MEB, GP, HV, LR, J-BM, and BH contributed to the code routines. TB is the main developer of the Genuage platform. BH provided super-resolution and single-particle tracking data. MEB, J-BM, and BH conceived the idea. J-BM and BH directed the project and wrote the paper. All authors read the paper and contributed to and approved the content.

FUNDING

We acknowledge funding from the Fondation pour la recherche médicale (FRM; DEI20151234398), the Agence National de la recherche (ANR-19-CE42-0003-01), the LabEx CelTisPhyBio (ANR-11-LABX-0038, ANR-10-IDEX-0001-02), and the Institut Curie. We recognize the support of France-BioImaging infrastructure grant ANR-10-INBS-04 (Investments for the future). We acknowledge the financial support of the Agence pour la Recherche sur le Cancer (ARC Foundation), DIM ELICIT and from ITMO Cancer of Aviesan on funds administered by Inserm (grant N° 20CP092-00). We acknowledge funding from the Institut Pasteur, the Institut Curie, the Paris Science lettre (PSL) University, the sponsorships of CRPCEN, Gilead Science and foundation EDF, the INCEPTION project (PIA/ANR-16-CONV-0005, OG), the programme d'investissement d'avenir supported by L'Agence Nationale de la Recherche ANR-19-P3IA-0001 Institut 3IA Prairie and the support of the AVIRON grant from the Région Ile-de-France (DIM-ELICIT).

ACKNOWLEDGMENTS

We acknowledge the naviscope Inria project lab teams and Thomas Isenberg for fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbinf.2021.775379/full#supplementary-material

REFERENCES

AnDi (2020) andi-challenge Available at: http://andi-challenge.org/.

Bergé, L.-P., Aouf, N., Duval, T., and Coppin, G. (20162016). "Generation and VR Visualization of 3D point Clouds for Drone Target Validation Assisted by an Operator," in *8th Computer Science and Electronic Engineering (CEEC)*, 66–70. doi:10.1109/CEEC.2016.7835890

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., et al. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* 313, 1642–1645. doi:10.1126/science.1127344

Blanc, T., El Beheiry, M., Caporal, C., Masson, J. B., and Hajj, B. (2020). Genuage: Visualize and Analyze Multidimensional Single-Molecule point Cloud Data in Virtual Reality. *Nat. Methods* 17, 1100–1102. doi:10.1038/s41592-020-0946-1

Borland, D., and Taylor, M. R. (2007). Rainbow Color Map (Still) Considered Harmful. *IEEE Comput. Graph Appl.* 27, 14–17. doi:10.1109/MCG.2007.323435

Brůža, V., Byška, J., Mičan, J., and Kozlíková, B. (2021). VRdeo: Creating Engaging Educational Material for Asynchronous Student-Teacher Exchange Using Virtual Reality. *Comput. Graphics* 98, 280–292. doi:10.1016/j.cag.2021.06.009

Charles, R. Q., Su, H., Kaichun, M., and Guibas, L. J. (2017). "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI (IEEE), 77–85. doi:10.1109/CVPR.2017.16

Chenouard, N., Smal, I., de Chaumont, F., Maška, M., Sbalzarini, I. F., Gong, Y., et al. (2014). Objective Comparison of Particle Tracking Methods. *Nat. Methods* 11, 281–289. doi:10.1038/nmeth.2808

Doutreligne, S., Cragnolini, T., Pasquali, S., Derreumaux, P., and Baaden, M. (2014). "UnityMol: Interactive Scientific Visualization for Integrative Biology," in 2014 IEEE 4th Symposium on Large Data Analysis and Visualization (LDAV), 109–110. doi:10.1109/LDAV.2014.7013213

El Beheiry, M., Dahan, M., and Masson, J. B. (2015). InferenceMAP: Mapping of Single-Molecule Dynamics with Bayesian Inference. *Nat. Methods* 12, 594–595. doi:10.1038/nmeth.3441

El Beheiry, M., and Dahan, M. (2013). ViSP: Representing Single-Particle Localizations in Three Dimensions. *Nat. Methods* 10, 689–690. doi:10.1038/nmeth.2566

El Beheiry, M., Doutreligne, S., Caporal, C., Ostertag, C., Dahan, M., and Masson, J. B. (2019). Virtual Reality: Beyond Visualization. *J. Mol. Biol.* 431, 1315–1321. doi:10.1016/j.jmb.2019.01.033

El Beheiry, M., Godard, C., Caporal, C., Marcon, V., Ostertag, C., Sliti, O., et al. (2020). DIVA: Natural Navigation inside 3D Images Using Virtual Reality. *J. Mol. Biol.* 432, 4745–4749. doi:10.1016/j.jmb.2020.05.026

Fey, M., and Lenssen, J. E. (2019). Fast Graph Representation Learning with PyTorch Geometric. arXiv:1903.02428 [cs, stat]. Available at: http://arxiv.org/abs/1903.02428.

Gaia Sky (2014). Zentrum für Astronomie. Available at: https://zah.uni-heidelberg.de/gaia/outreach/gaiasky.

Genuage GitHub (2020). Genuage GitHub. Available at: https://github.com/genuage/genuage.

Grottel, S., Krone, M., Müller, C., Reina, G., and Ertl, T. (2015). MegaMol--A Prototyping Framework for Particle-Based Visualization. *IEEE Trans. Vis. Comput. Graph* 21, 201–214. doi:10.1109/TVCG.2014.2350479

Hajj, B., Wisniewski, J., El Beheiry, M., Chen, J., Revyakin, A., Wu, C., et al. (2014). Whole-cell, Multicolor Superresolution Imaging Using Volumetric Multifocus Microscopy. *Proc. Natl. Acad. Sci. U S A.* 111, 17480–17485. doi:10.1073/pnas.1412396111

Höfling, F., and Franosch, T. (2013). Anomalous Transport in the Crowded World of Biological Cells. *Rep. Prog. Phys.* 76, 046602. doi:10.1088/0034-4885/76/4/046602

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., et al. (2008). Robust Single-Particle Tracking in Live-Cell Time-Lapse Sequences. *Nat. Methods* 5, 695–702. doi:10.1038/nmeth.1237

Lichtman, J. W., and Conchello, J. A. (2005). Fluorescence Microscopy. *Nat. Methods* 2, 910–919. doi:10.1038/nmeth817

Liu, W., Sun, J., Li, W., Hu, T., and Wang, P. (2019). Deep Learning on Point Clouds and its Application: A Survey. *Sensors (Basel)* 19, 4188. doi:10.3390/s19194188

Liu, W., Lai, B., Wang, C., Bian, X., Yang, W., Xia, Y., et al. (2020). "Learning to Match 2D Images and 3D LiDAR Point Clouds for Outdoor Augmented Reality," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 654–655. doi:10.1109/VRW50115.2020.00178

MacQueen, J. (1967). "Some Methods for Classification and Analysis of Multivariate Observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Statistics), 1, 281–297.

Manzo, C., and Garcia-Parajo, M. F. (2015). A Review of Progress in Single Particle Tracking: from Methods to Biophysical Insights. *Rep. Prog. Phys.* 78, 124601. doi:10.1088/0034-4885/78/12/124601

Masson, J. B., Dionne, P., Salvatico, C., Renner, M., Specht, C. G., Triller, A., et al. (2014). Mapping the Energy and Diffusion Landscapes of Membrane Proteins at the Cell Surface Using High-Density Single-Molecule Imaging and Bayesian Inference: Application to the Multiscale Dynamics of Glycine Receptors in the Neuronal Membrane. *Biophys. J.* 106, 74–83. doi:10.1016/j.bpj.2013.10.027

Matthews, D. (2018). Virtual-reality Applications Give Science a New Dimension. *Nature* 557, 127–128. doi:10.1038/d41586-018-04997-2

Miller, H., Zhou, Z., Shepherd, J., Wollman, A. J. M., and Leake, M. C. (2017). Single-molecule Techniques in Biophysics: a Review of the Progress in Methods and Applications. *Rep. Prog. Phys.* 81, 024601. doi:10.1088/1361-6633/aa8a02

Muñoz-Gil, G., Volpe, G., García-March, M. A., Metzler, R., Lewenstein, M., and Manzo, C. (2020). "The Anomalous Diffusion challenge: Single Trajectory Characterisation as a Competition," in *Emerging Topics in Artificial Intelligence 2020* (International Society for Optics and Photonics), 114691C. doi:10.1117/12.2567914

Nehme, E., Freedman, D., Gordon, R., Ferdman, B., Weiss, L. E., Alalouf, O., et al. (2020). DeepSTORM3D: Dense 3D Localization Microscopy and PSF Design by Deep Learning. *Nat. Methods* 17, 734–740. doi:10.1038/s41592-020-0853-5

Pavez, E., Chou, P. A., de Queiroz, R. L., and Ortega, A. (2018). Dynamic Polygon Clouds: Representation and Compression for VR/AR. *APSIPA Trans. Signal Inf. Process.* 7. doi:10.1017/ATSIP.2018.15

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. arXiv:1706.02413 [cs]. Available at: http://arxiv.org/abs/1706.02413.

Ramirez, P. Z., Paternesi, C., Luigi, L. D., Lella, L., Gregorio, D. D., and Stefano, L. D. (2020). "Shooting Labels: 3D Semantic Labeling by Virtual Reality," in 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 99–106. doi:10.1109/AIVR50618.2020.00027

Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* 3, 793–795. doi:10.1038/nmeth929

Spark, A., Kitching, A., Esteban-Ferrer, D., Handa, A., Carr, A. R., Needham, L. M., et al. (2020). vLUME: 3D Virtual Reality for Single-Molecule Localization Microscopy. *Nat. Methods* 17, 1097–1099. doi:10.1038/s41592-020-0962-1

Stets, J. D., Sun, Y., Corning, W., and Greenwald, S. W. (2017). "Visualization and Labeling of point Clouds in Virtual Reality," in *SIGGRAPH Asia 2017 Posters SA 17* (New York, NY, USA: Association for Computing Machinery), 1–2. doi:10.1145/3145690.3145729

Verdier, H., Duval, M., Laurent, F., Casse, A., Vestergaard, C. L., Masson, J. B., et al. (2021). Learning physical properties of anomalous random walks using graph neural networks. *J. Phys. A Math. Theoret.* 54(23), 234001. doi:10.1038/s41592-020-0962-1

Voxel Learning in DIVA (2020). Voxel Learning in DIVA. Available at: https://github.com/DecBayComp/VoxelLearning.

Wirth, F., Quehl, J., Ota, J., and Stiller, C. (2019). "PointAtMe: Efficient 3D Point Cloud Labeling in Virtual Reality," in 2019 IEEE Intelligent Vehicles Symposium (IV), 1693–1698. doi:10.1109/IVS.2019.8814115

Zhang, C. (2020). "Investigation on Motion Sickness in Virtual Reality Environment from the Perspective of User Experience," in 2020, IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), 393–396. doi:10.1109/ICISCAE51034.2020.9236907

Conflict of Interest: MEB and J-BM are cofounders, shareholders and, respectively, Chief Technology Officer (CTO) and Chief Scientific Officer (CSO) of AVATAR MEDICAL SAS, a startup that commercialises software for surgery planning in virtual reality. AVATAR MEDICAL had no role in study design, data collection and analysis, funding, decision to publish, or preparation of the manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers

Supplementary Material

1 TCP/IP Communication between Genuage and Python or Matlab processes

Systems using the ZeroMQ framework have been implemented in order to establish communication between Genuage and other processes using the TCP/IP protocol. Thus, it is possible for Genuage to send and receive byte data, Python and Matlab Interfaces are provided. This functionality can be used in two implementations, the first allows users to send numeric data from a MATLAB or python process to Genuage to generate a new cloud, the second is used in the real time inference module, where back and forth communications between Genuage and a python process running a neural network are used to send data for the neural network to infer upon, and get the results back for Genuage to display and save.

2 User-defined DLLs

We provide several DLL examples that are compatible with Genuage to illustrate the different possible input and output formats when performing point cloud quantifications.

The codes are based on Matlab scripts that were transformed into DLLs. The original Matlab files are also provided in the GitHub repository.

The provided DLL are for analyzing single-molecule point-cloud data sets.

- *Density*: The purpose of this DLL is to estimate for each point the number of the neighbors within a given 3D radius. The user inputs the desired radius. The DLL upgrades the raw data set with a new column of the calculated density values.
- *CorrectBlinking*: Analyzing single-molecule localization data sets requires specific considerations especially for super-resolution microscopy data sets. A common artifact arises from the blinking of single molecules. The same molecule can last on multiple frames and thus appears as multiple detected points. The reconstructed image can thus conceal the real underlying molecular distribution. Points arising from the same molecule need to accounted for and combined into single detection.
  This DLL is dedicated for the correction of multiple appearances due to blinking of the same molecule in PALM and STORM imaging. The DLL combines the localizations that are found at a given spatial and temporal proximity. The exact temporal and spatial extents can be estimated by several established methods (Betzig et al., 2006; Rust et al., 2006; Annibale et al., 2011; Coltharp et al., 2012; Puchner et al., 2013; Bohrer et al., 2021). The values are entered manually as numerical input when executing the DLL.
  The code loops over all the frames of the input point-cloud dataset. For each frame, a second loop is performed over a number of frames within search time window (e.g. over 20 frames if the search time is defined at 1 s and the acquisition frame rate is 50 ms). The distances between all the molecules of the considered frame and all the molecules of the following frames are computed. Localizations that can be considered as coming from a same blinking molecule are identified and labelled with an identical index. At the end, the algorithm returns for each localization of each frame its index corresponding to a given unique molecule. Based on this information, a mean position and a mean frame number is computed for each multiple-

appearing molecule. This DLL generates finally a new point cloud with a different number of elements compared to the original one.

- *ApparentDiffusion*: The DLL is dedicated for dynamic point cloud recordings that are generated from single particle tracking experiments. The diffusion of the particles is highly dependent on its physical characteristics (size and functionalization) as well as the surrounding media. As a first order approximation, apparent diffusion coefficient of each trajectory is retrieved from the mean-square displacement (MSD) at different time steps (Normanno et al., 2015; Nora et al., 2020). The apparent diffusion coefficient is estimated from the MSD values that are computed between 2dt$2dt$ and 5dtdt, dt being the exposure time of each acquisition frame. The DLL estimates the diffusion coefficient D of a selected trajectory by a linear fitting of the MSD. The slope is related to the apparent diffusion coefficient by: $slope = 2 \times N \times D$. Here N stands for the dimensionality of the data, e.g. 2 for $2D$ and 3 for $3D$ recordings. The DLL returns the estimated diffusion coefficient of the selected trajectory based on the dimensionality of the data.

We note here that manual inputs are still necessary in some analysis routines, for instance to define the research radius in *Density* and dimensionality in *ApparentDiffusion*. We designed a general interface for providing the necessary numeric arguments (of various types: integers, floats and doubles) for the DLLs.

3 Calculating and presenting physical property map

The 3D maps are generated in a 3-step process.

First, a K-means Clustering algorithm is used to partition the points into a number of clusters based on proximity. The K-means algorithm iterates over a predefined set of starting clusters. Users can set the number of clusters or partitions, as well as the maximum number of iterations for the algorithm to perform. A new data column is added to the cloud at this step, each entry being the id number of the cluster the point belongs to.

Second, Bayesian inference is performed on each cluster of points, using the inference algorithm already included in Genuage (Blanc et al., 2020). Theoretical bases of the inference are described in (Masson et al., 2014; Beheiry et al., 2015). A diffusion coefficient is thus calculated for all the dynamic points within each cluster. User can set various parameters for the priors and the level of noise to be considered by the algorithm. Similarly, a diffusion force vector can be estimated. A new data column is later added to the cloud containing the corresponding diffusion coefficient.

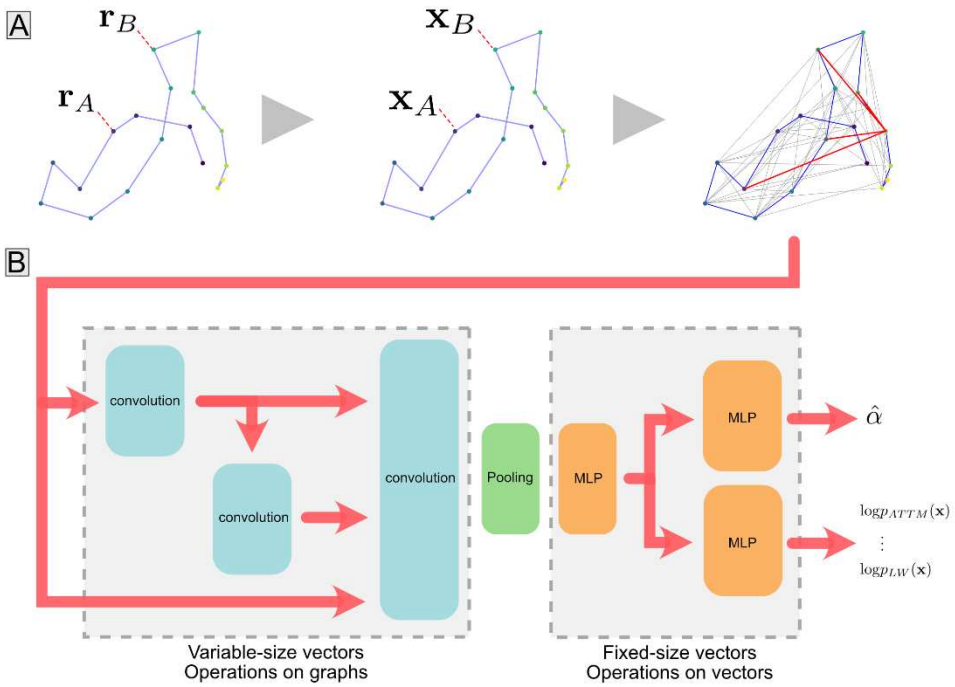
Finally, an implementation of the Quickhull (Barber et al., 1996) algorithm is used to create a 3-dimensional convex hull mesh for the points in each cluster. The resulting meshes can be displayed as solids, colored according to the values of the diffusion coefficient mapped to one of Genuage's colormaps, or as wireframes. The visibility and opacity of rendered meshes and surfaces can be controlled by the user. The force vectors for the diffusion within each cluster, is displayed as an arrow with a size proportional to the force vector's norm.

4 3D Map Rendering peculiarities

The meshes for the polygons generated by the 3D physical map module are generated with split vertices, so that no two triangles share the same vertex. This allows for the edges of each face of the polygons to be lit separately by Unity's lighting engine and for the edges of the solids to be clearly

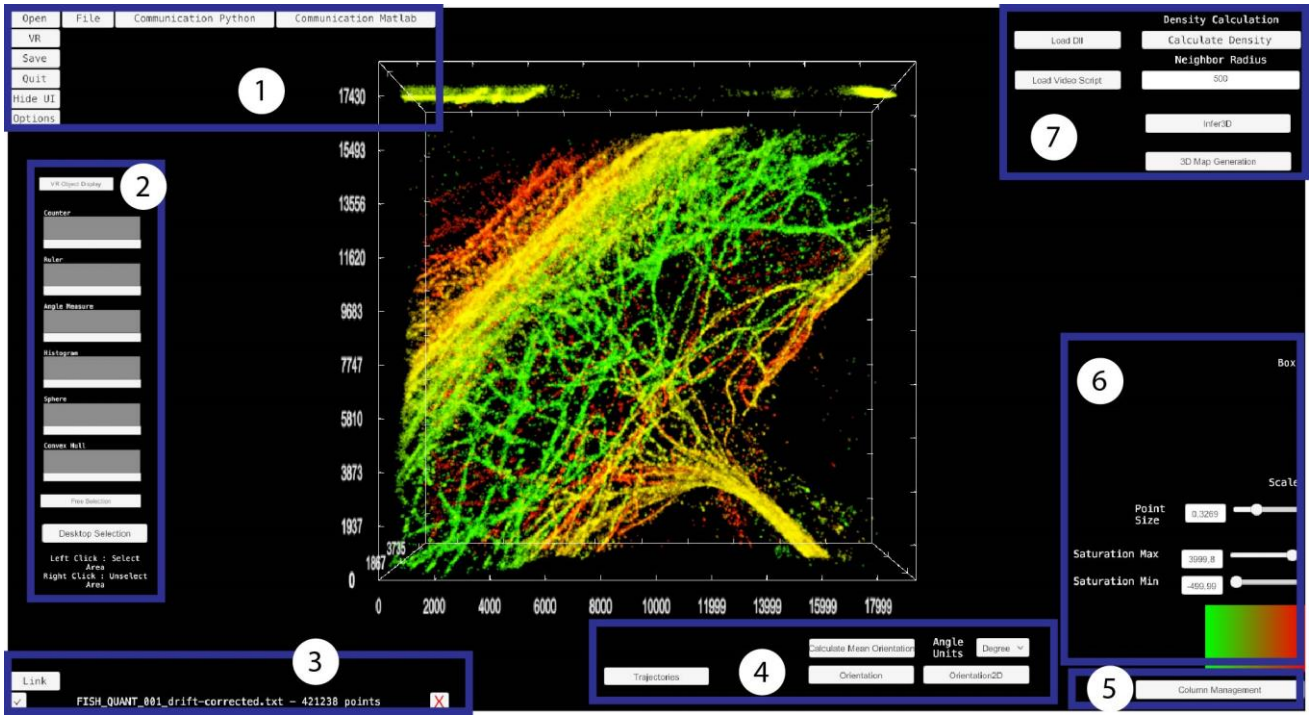
visible when rendered. The meshes are rendered with a modified version of Unity's standard surface shader, in order to take advantage of Unity's lighting system and be compatible with Genuage's shader-based interactions like the VR Clipping Plane tool.

5 Pretrained neural network

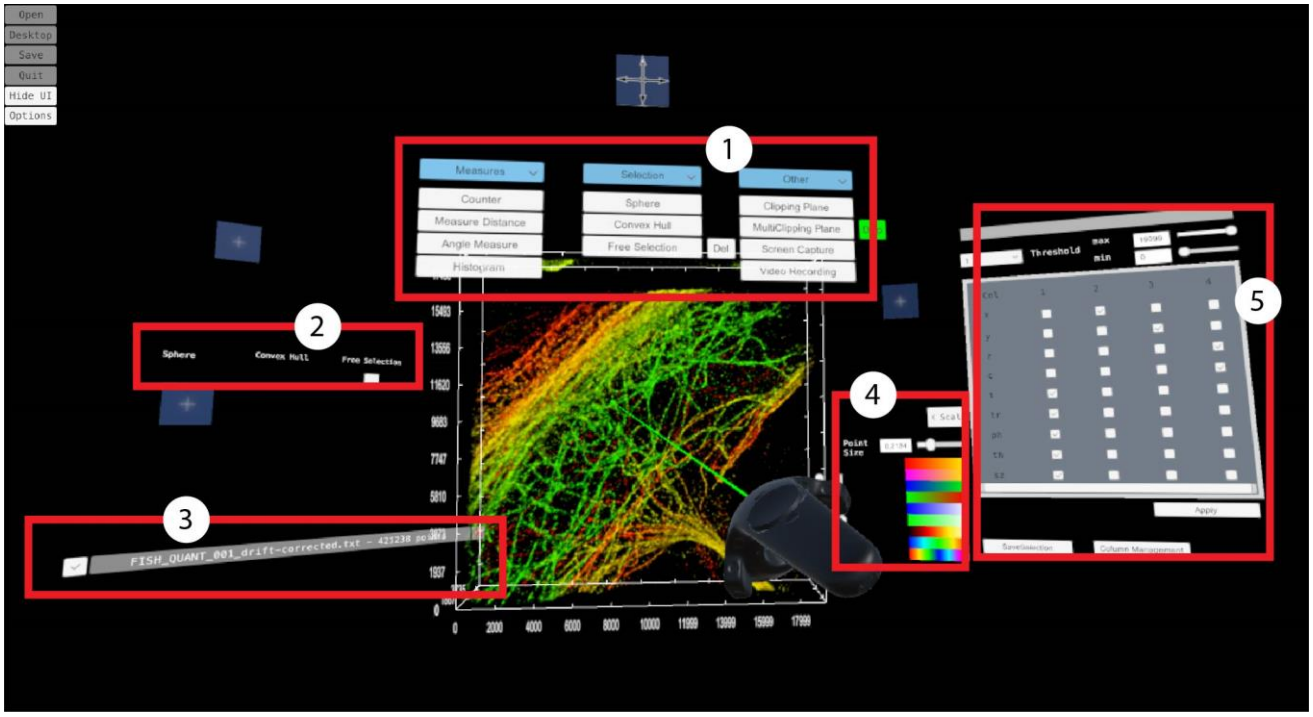


Supplementary Figure 1: a) Construction of a graph associated to a trajectory: a graph is built from successive spatial positions by considering locations (the point cloud) as nodes and by drawing edges between nodes (at random or following a deterministic pattern). Feature vectors are then computed for each node (position, distance to origin, time coordinate, etc.) and for each edge (relative displacement, mean step size of the trajectory between both positions, time difference, etc.). b) Schematic view of a convolution operation: given a "receiving" node (here in red), messages are computed for each neighboring node, from the vector resulting of the concatenation of the neighbor node's features and of the edge features. This vector is given as input to a trainable multi-layer-perceptron (MLP), whose output is a message. After the convolution, each node's feature vector is updated using an aggregation (feature average or maximum) of all the messages it received. The Model architecture: three depths of convolutions are successively concatenated. After the last convolution, features of all nodes are averaged, with an attention mechanism (Knyazev et al., 2019). This averaged feature vector is a latent representation of the trajectory, from which the property of interest (here, the anomalous exponent) is inferred using a classic MLP.

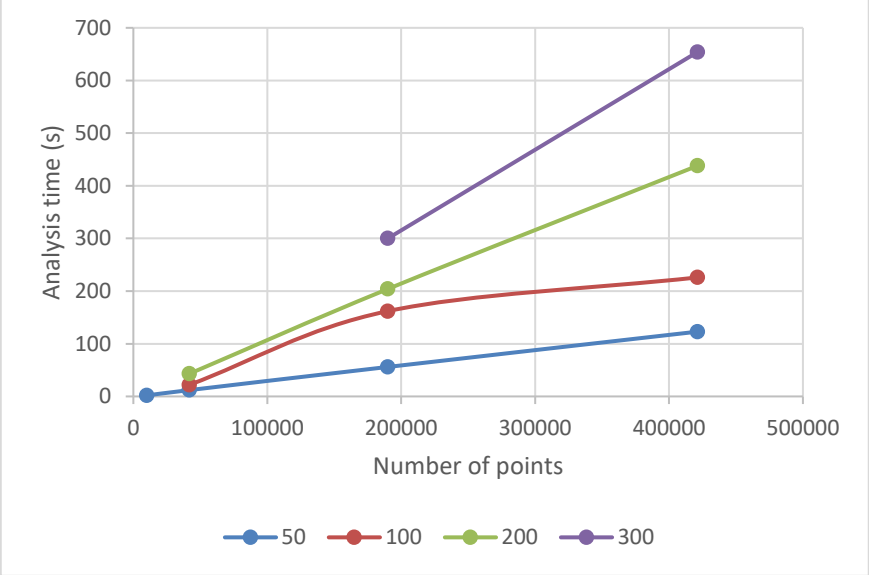
6 Supplementary Figures



Supplementary Figure 2: General overview of the interface of Genuage in desktop mode. 1: File import and export, communication with external software. 2: objects and selections interface. 3: opened files and multichannel handling. 4: advanced multidimensional data visualization specifically conceived for dynamic data and molecular orientation representation. 5: column management and thresholding. 6: visual parameters control. 7: advanced point cloud analysis.



Supplementary Figure 3: General overview of the interface of Genuage in VR mode. 1: A set of tools dedicated to interact and analyze point cloud in the VR mode. It includes (i) performing VR facilitated measurements such as counting, distance, angle and histogram measurements, (ii) specific 3D selections and multi-selections, and (iii) visualization tools such as clipping plane and video recording. 2: Selections and objects interface. 3: Channel interface. 4: Visualization parameters adjustments. 5: Column configuration and thresholding.



Supplementary Figure 4: Performance characterization of the 3D inference map algorithm as a function of the number of points, evaluated for several number of clusters (50, 100, 200 and 300).

7 Supplementary table

Table comparing different software platforms dedicated to visualizing and analyzing point cloud data. The only relevant comparaison for single-molecule data is between VISP, vLUME and Genuage.

Functionalities		**VISP**	**PointCloud XR**	**vLUME**	**CellexalVR**	**Genuage**
Interface	Desktop mode	✓	×	×	×	✓
	VR mode	×	✓	✓	✓ yes	✓
	Opening multiple files	✓ into individual tabs	×	✓ in the same working space	×	✓ into individual tabs
	Accepted format	Text files	LAS 1.2 format	csv	✓ Seurat or Scanpy data converted by a custom R package.	Text files
VR compatibility	Headsets		HTC Vive, HTC Vive Pro	HTC Vive, HTC Vive Pro Oculus Rift and Rift S / Quest	HTC Vive, Valve Index	HTC Vive, HTC Vive Pro, Oculus Rift and Rift S / Quest
Visual adjustment	Color adjustment	✓	✓	✓	✓	✓
	Axis scaling	×	×	✓	✓	✓
	Point cloud size	✓	✓	✓	✓	✓
	Point cloud size by localization precision	✓	×	×		✓
	Multi-color channels	✓	×	✓	✓	✓

Data input and visualization parameters	Modular column assignment	×	×	×	×	✓
	Color code assigned to specific column	×	✓	✓	✓	✓
	3D positions	✓	✓	✓	✓	✓
	Trajectories	×	×	✓	✓	✓
	Showing in time	×	✓	✓	✓	✓
	2D Orientation	×	×	×	×	✓
	3D Orientation	×	×	×	×	✓
	Thresholding	✓	×	✓	✓	✓
VR tools	Clipping plane	-	×	✓	×	✓
	Freezable Clipping plane	-	×	×	×	✓
	Selections	-	✓	✓	✓	✓
	Counting	-	✓	×	✓	✓
	Measuring distances	✓in desktop mode	✓	×	×	✓
	Histogram profiler	✓in desktop mode	×	×	×	✓
	Annotation in VR	-	×	✓	✓	×
	Data manipulation	-	×	✓	✓	×

	in VR (displacement)					
Built in analysis tools	Point cloud local density calculation	✓	×	✓	×	✓
	Ripley's K function	×	×	✓	×	×
	Localization density (in a selection)	×	×	✓	×	×
	Largest and shortest distance	×	×	✓	×	×
	Diffusion coefficient	×	×	×	×	✓ $
	Drift	×	×	×	×	✓ $
	Custom analyses by user made algorithms	×	×	✓through C# scripts	×	✓through C++/Matlab/Other based DLLs
	3D physical properties map generation	×	×	×	×	✓
Reading/writing	Exporting videos and figures	✓embeded	✓ Through external software	✓embeded	✓Exporting Images and html reports	✓ embeded
	Saving progress	×	✓	✓	✓ Various saving features	✓ Saving JSON file
	Exporting data	✓	×	✓	✓	✓ Exporting JSON file, exporting selections

	Communication with external software	×	×	×	✓ Communication with R servers	✓

$ Genuage provides up to 9 different methods to calculate the diffusion coefficient, and 10 additional methods to calculate diffusion and drift.

8 Supplementary Videos

Supplementary Video 1: Live dynamic inference performed in Genuage using a pretrained network in Python. The VR experience is not affected by the calculation.
https://youtu.be/ZkG_D0sQlLA

Supplementary Video 2: Examples illustrating the different analysis and possible cloud manipulations in Genuage when running user-defined DLLs.
https://youtu.be/J-WrA3YwJ-8

Supplementary Video 2 bis : A show case for DLL application on a selected trajectory in VR
https://youtu.be/U3bTAy1IVko

Supplementary Video 3: Video recording in Genuage.
https://youtu.be/MWk1Nsg0glA

Supplementary Video 4: Principle of generating 3D diffusion maps in Genuage.
https://youtu.be/VrXVMyuAuCo

9 Supplementary references

Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U., and Radenovic, A. (2011). Quantitative Photo Activated Localization Microscopy: Unraveling the Effects of Photoblinking. *PLOS ONE* 6, e22678. doi:10.1371/journal.pone.0022678.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* 22, 469–483. doi:10.1145/235815.235821.

Beheiry, M. E., Dahan, M., and Masson, J.-B. (2015). InferenceMAP: mapping of single-molecule dynamics with Bayesian inference. *Nat Methods* 12, 594–595. doi:10.1038/nmeth.3441.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., et al. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313, 1642–1645.

Blanc, T., El Beheiry, M., Caporal, C., Masson, J.-B., and Hajj, B. (2020). Genuage: visualize and analyze multidimensional single-molecule point cloud data in virtual reality. *Nat Methods* 17, 1100–1102. doi:10.1038/s41592-020-0946-1.

Bohrer, C. H., Yang, X., Thakur, S., Weng, X., Tenner, B., McQuillen, R., et al. (2021). A pairwise distance distribution correction (DDC) algorithm to eliminate blinking-caused artifacts in SMLM. *Nat Methods* 18, 669–677. doi:10.1038/s41592-021-01154-y.

Coltharp, C., Kessler, R. P., and Xiao, J. (2012). Accurate Construction of Photoactivated Localization Microscopy (PALM) Images for Quantitative Measurements. *PLOS ONE* 7, e51725. doi:10.1371/journal.pone.0051725.

Knyazev, B., Taylor, G. W., and Amer, M. R. (2019). Understanding Attention and Generalization in Graph Neural Networks. *arXiv:1905.02850 [cs, stat]*. Available at: http://arxiv.org/abs/1905.02850 [Accessed September 9, 2021].

Masson, J.-B., Dionne, P., Salvatico, C., Renner, M., Specht, C. G., Triller, A., et al. (2014). Mapping the Energy and Diffusion Landscapes of Membrane Proteins at the Cell Surface Using High-Density Single-Molecule Imaging and Bayesian Inference: Application to the Multiscale Dynamics of Glycine Receptors in the Neuronal Membrane. *Biophysical Journal* 106, 74–83. doi:10.1016/j.bpj.2013.10.027.

Nora, E. P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., et al. (2020). Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* 11, 5612. doi:10.1038/s41467-020-19283-x.

Normanno, D., Boudarène, L., Dugast-Darzacq, C., Chen, J., Richter, C., Proux, F., et al. (2015). Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun* 6, 7357. doi:10.1038/ncomms8357.

Puchner, E. M., Walter, J. M., Kasper, R., Huang, B., and Lim, W. A. (2013). Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory. *PNAS* 110, 16015–16020. doi:10.1073/pnas.1309676110.

Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods* 3, 793–796.

Perspectives

Improving detection of learning

The future approach to modeling our data will attempt to reduce all the steps leading from raw images to latent representations. The first approach we are considering is the construction of a variational autoencoder applied directly to the videos of the series of experiments, *i.e.* (512X512X300) to avoid manually defining ROIs. Here, this would amount to encoding matrices of size (512X512X300). A 3D U-Net[1,2] type architecture will be considered first where time will be considered as a third, spatial dimension. This auto-encoding space will be trained on many different mice and on different protocols in order to create a better understanding. In addition, special attention will be paid to the data augmentation procedure with numerous spatial morphologic transformations. This will ensure that any point in space has been sampled with different types of neural geometries. Focus will also be put on temporal transformations that will include a priori knowledge of the expected signal structures to make the representation robust to temporal translations and non-specific activity noise. Finally, this autoencoder will be initialized by a self-supervised learning task. It will consist of a spatial and temporal inpainting task. This will allow both learning of the expected local and global structures of the interneurons and their temporal dynamics.

In addition, an evolution of the protocol is envisaged in order to make the components resulting from learning more evident. Our results have shown that the decoding of odors both in terms of their nature and their proportion is possible even in a very sparse population of neurons, so the protocol can be reduced to two fixed odors for the whole experimental series. In order to capture more of the learning kinetics in the neural population, a 4-arm experiment will probably be needed. This would involve (i) a classical learning arm with the association of the correct choice with a reward, (ii) an arm with dissociation between reward and choice, (iii) an arm where until the middle of the duration there is the association between choice and reward, then independence between the two and finally (iv) an arm where until the middle of the duration there is the dissociation between reward and choice then the association between reward and choice.

This approach should allow for a more refined analysis of the representation of learning in the sampled neural population.

From architecture to function

Finding a statistical signature of learning provides a pathway for future studies to determine which activity patterns govern specific behavioral states. Our autoencoder results will be further deciphered to provide circuit level associations that can be utilized to develop a computational model to replicate the learning process. Past and current work in the Lledo lab, in collaboration with Hermann Riecke at Northwestern University, has led to a structural plasticity model that was able to predict greater odor contrast enhancement, wholly based on experimental structural data[10]. Now, with the vast amount of Ca^{2+} imaging data, we can refine this model to produce a spiking model that can test the predictions of the autoencoder and aid in refining the structure of the model for providing biologically relevant associations.

The OB is not an isolated unit but receives a high degree of various top-down inputs that modulate its output to the cortex[11]. The results from our autoencoder can guide us to which modulatory inputs could be influencing GC activity changes associated with learning. We plan to perform future experiments that interrupt certain components of the circuit to determine the effect on behavior and whether the classifier can still predict the activity changes associated with the behavior. This would drive toward a causal, mechanistic understanding of the behavior circuits which could lead to a more complete circuit map of the OB and could potentially be applied to other primary sensory systems.

Linking structure and function is a fundamental pursuit in neuroscience since the connectivity and associated activity define the functional outputs ultimately leading to behavior[12]. In most other organ systems, the predominant function of the organ is dependent on repeatable units with the same defined function. For example, in the kidney, knowing the function of a single nephron makes it possible to approximate the full function of the organ. In stark contrast, the vast and complex connectivity of the CNS

predominantly defines its function. In the past two decades there has been a tremendous effort to accurately and rapidly segment neurons from either fluorescent or electron microscopy images. Combining 3D images of sparsely labeled GCs with our self-supervised segmentation method, we aim to make a high-resolution circuit map of the OB which can feed our computational model. This will allow us to include cable theory, spine morphology and other structural parameters that have yet been overlooked, and which will likely provide greater accuracy of the model in replicating the biological system. Combining unbiased methods for finding correlations in activity with behavior and an accurate map of the OB circuit can also provide a structure-function framework that can be applied to other brain regions.

Self-supervised learning and virtual reality extraction of neural tree architectures

In order to improve the performance of approaches combining manual intervention, virtual reality, and machine learning in the context of neural tree segmentation, a promising direction is self-supervised learning[3–5] (SSL). SSL consists of exploiting unannotated databases in order to extract information from them before exploiting the accessible annotated data, ultimately leading to more powerful machine learning models. The study of unannotated databases is done through pretextual tasks, in which a problem is posed on the data that requires only the data itself to answer. Examples of predicate tasks are inpainting where a neural network has to "fill in" holes made in images, or predicting whether two data sets are neighbors either in an image or in a time series. These tasks allow for a learning process of data structures that can then be exploited for supervised tasks.

Two SSL approaches will help to improve the segmentation of neural trees on small data sets. The first one is intended to be included directly in DIVA-cloud[6] in order to perform voxel learning. In this procedure, we seek to associate to each voxel an informative feature vector in order to learn from some annotations made on a single dataset. SSL will here take the form of inpainting tasks[7–9] done on small 3D patches where the pretext task consists of either predicting the interior or exterior of the patch. The

latent space learned in this task can then be associated with the set of voxels and used as local features. We show an example of results of such an approach applied to MRI imaging of a fetus in Fig. 18 from a recent submission to the MICCAI 2022 conference.

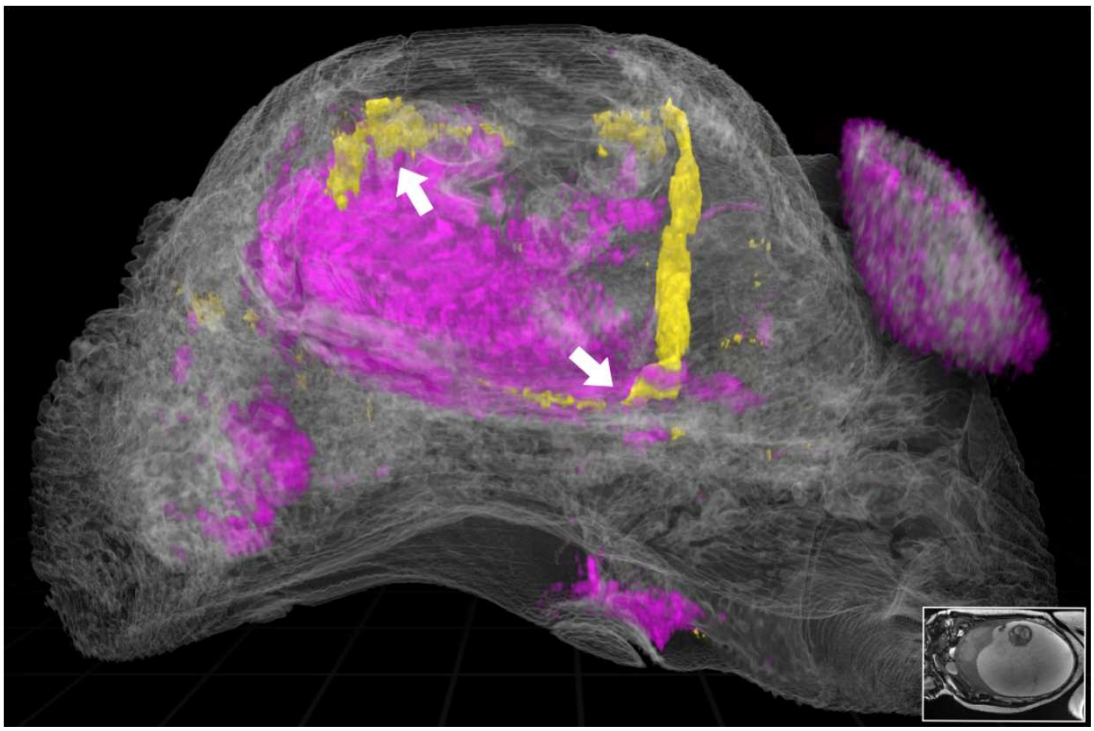


Figure 18: Fused representation of umbilical cord segmentation (in yellow), placenta segmentation (in magenta) and raw data in gray. A representation as a z-stack is shown in the bottom right corner. The insertions of the cords in the placenta are denoted with white arrows.

The second approach is global. Neural tree images are characterized by their sparsity. Neurons are tubular structures, where their main direction can vary within the recorded volume. A natural pretext task is to shuffle the ordering of the recorded slices (in various directions) and ask the neural network to reorder them. This SSL approach will be applied after the application of DIVA-cloud to segment entire datasets from the local diva-cloud tagging plus final manual proofing.

References

1. Khawam, M. *et al.* Fetal Brain Biometric Measurements on 3D Super-Resolution Reconstructed T2-Weighted MRI: An Intra- and Inter-observer Agreement Study. *Front. Pediatr.* **9**, 651 (2021).
2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *ArXiv160606650 Cs* (2016).
3. Chen, L. *et al.* Self-supervised learning for medical image analysis using image context restoration. *Med. Image Anal.* **58**, 101539 (2019).
4. Shurrab, S. & Duwairi, R. Self-supervised learning methods and applications in medical imaging analysis: A survey. *ArXiv210908685 Cs Eess* (2021).
5. Klinghoffer, T. *et al.* Self-Supervised Feature Extraction for 3D Axon Segmentation. *ArXiv200409629 Cs Eess* (2020).
6. Corentin Guerinot et al. New approach to accelerated image annotation by leveraging virtual reality and cloud computing. *Front. Bioinforma.*
7. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context Encoders: Feature Learning by Inpainting. *ArXiv160407379 Cs* (2016).
8. Demir, U. & Unal, G. Patch-Based Image Inpainting with Generative Adversarial Networks. *ArXiv180307422 Cs* (2018).
9. Yan, Z., Li, X., Li, M., Zuo, W. & Shan, S. Shift-Net: Image Inpainting via Deep Feature Rearrangement. *ArXiv180109392 Cs* (2018).
10. Sailor, K. A. *et al.* Persistent Structural Plasticity Optimizes Sensory Information Processing in the Olfactory Bulb. *Neuron* **91**, 384–396 (2016).
11. Wen, P. *et al.* Cortical Organization of Centrifugal Afferents to the Olfactory Bulb: Mono- and Trans-synaptic Tracing with Recombinant Neurotropic Viral Tracers. *Neurosci. Bull.* **35**, 709–723 (2019).
12. Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).

Segmentation on-the-fly of fetal MRI with virtual reality and one-shot learning

Charlotte Godard[1,2], Corentin Guérinot[1,3,4], Juliette Virfollet[5], Charline Henry[5], Raphaël Bartin[5], Chloé Arthuis[5], David Grévent[5], Laurence Bussières[5], Laurent Salomon[5], and Jean-Baptiste Masson[1]

[1] Decision and Bayesian Computation, USR 3756 (C3BI/DBC) & CNRS UMR 3751, Institut Pasteur, Paris, France
[2] École doctorale Physique en Île-de-France, PSL University, Paris, France
[3] Perception and Memory Unit, CNRS UMR3571, Institut Pasteur, Paris, France
[4] Sorbonne Université, Collège doctoral, Paris, France
[5] LUMIERE, Hôpital universitaire Necker-Enfants Malades, Université de Paris, France
jbmasson@pasteur.fr

Abstract. All medical imaging data are not standardized. Thus, fetal nuclear magnetic resonance imaging (MRI) is highly variable due to the maternal morphology and fetus positional dynamics. Fetal surgery planning requires complex segmentation on images where some organs may not be easily identified. We introduce an approach combining virtual reality (VR), one-shot learning, and a few user's gestures to enable rapid segmentation. Associating ray-casting with a dynamic transfer function generation process, the user can modify the appearance of the data to accelerate annotation on a whole volumetric representation. Based on few user tags combined to a set of multiple self-supervised, random, and handcrafted learned features, the entire volume is segmented through pixel classification. The computation is handled on an external server synchronized to the VR platform to ensure a smooth VR experience. Data fusion in VR enables simultaneous visualization of the original and segmented data, allowing possible correction in an iterative procedure. We present an application on the segmentation of placenta and umbilical cords in monochorionic twin pregnancies complicated by twin-twin transfusion syndrome (TTTS) from MRI images.

Keywords: Fetal MRI · Virtual Reality · Segmentation.

1 Introduction

Antenatal screening aims at improving perinatal care by identifying morphological abnormalities or prenatal diseases that may require specific management. It relies mainly on ultrasound (US) as the first-tier imaging modality. Prenatal US is primarily carried out in two dimensions with the evaluation of the different organs' presence, shape, and global morphology. Prenatal US scanning for addressing abnormal development lacks robustness, repeatability, and reproducibility.

The primary sources of complexities are lack of standardization of the analysis, difficult technical conditions, mother health variability, "geometry", and differences in the medical doctor's skills. Magnetic resonance imaging (MRI) is another imaging modality that could strongly reinforce robustness, reproducibility and significantly reduce user-dependent results. Modern MRIs provide high-quality images in a short amount of time, reducing motion artifacts. Finally, MRI is already used for prenatal diagnosis. The two imaging modalities, US and MRI, are complementary and could clarify diagnosis, postnatal prognosis and optimize perinatal management.

Multifetal gestation includes twins and higher-order multiples. Twins are increasing worldwide, presenting multiple challenges in management for the obstetrician. They account for a large proportion of neonatal morbidity and mortality. More specifically, monochorionic twins (where both fetuses share only one placenta) are subjected to specific complications, originating in either imbalance or abnormality of the placenta. This unequal placental sharing can cause complications including twin-twin transfusion syndrome (TTTS), twin anemia-polycythemia sequence (TAPS), selective intrauterine growth restriction or twin reversed arterial perfusion sequence (TRAP). Although the symptoms of all these complications are very different, the keystone of their management can be two-fold. When aiming at dual survival, it involves a surgical destruction of the inter-twin anastomoses on the chorionic plate. As for selective survival, it implies a permanent occlusion of the severely affected twin's cord to protect the normal co-twin. Both treatment require optimal imaging technologies for the best possible planing of this highly precise surgical procedure.

In the following we introduce a procedure to perform surgery planning of TTTS pregnancies based on segmentation on-the-fly performed in Virtual Reality (VR). We introduce a VR software platform to visualize and interact with raw MRI images featuring a procedure allowing users to adapt transfer function to local organ features. We present an accelerated system to extract features from the images, an annotation tool within the VR environment and a one-shot learning procedure to segment regions of interest. We compare these segmentations to data manually segmented by an expert and discuss future of surgery planning within VR environment.

2 Virtual reality and medical data

Although the technology and the underlying principles have not changed much since the 60's, VR has experienced a recent revival. An easy access to headsets, computers with powerful graphics cards and the emergence of libraries allowing easy and fast development (*i.e.* Unity, Unreal Engine, OpenXR) have enabled large scale applications redesign. Medical imaging by MRI and CT-scan produce three-dimensional data. They are however analyzed and explored by both radiologists and surgeons as two-dimensional stacks. VR offers the possibility to explore these data in 3D. Applications in medicine are mainly focused on surgery-specific topics and in education. Training initiatives have been devel-

oped in neurosurgical training [3], spinal surgery [11], orthopedic surgery [18] and patient education [6]. Clinical results are beginning to accumulate with applications heart diseases [13], breast cancer [16], liver surgery [10, 12], pediatric surgery [14] and orthopaedic surgery [17, 19].

However, VR representations often require prior processing such as segmentation. Hence, they are highly dependent on the quality of this processing, as segmentation error might be propagated into its VR representation. The ability to directly represent the raw data in VR is thus necessary to various image quality and modality. In addition, the user must be able to change the representation parameters in order to improve the visualization of a particular organ.

We have recently introduced the DIVA software platform [7] for the visualization and interaction in VR of any 3D imagery. The platform does not require any data pre-processing in order to build the VR representation. The VR model is obtained by a simple ray casting optimized to ensure a refresh rate of 80Hz. The user can adjust in real-time the transfer function, *i.e.* a look-up table allowing the adjustment of the voxel representation parameters according to its local characteristics. The platform relies on a dual-interface. The desktop mode allows to view the data in an interface imitating a radiologist's. It also provides a 3D representation on screen where the appearance of the data can be previewed and modified. The VR mode allows data exploration by navigating and slicing in the volume, in order to better visualize a target organ or possibly to put landmarks in it [7].

The analysis of MRI data presents particular challenges when compared to CT-scan images. They have poorer contrast between organs, are more sensitive to patient movement and are generally more variable between patients and machines. Thus, the ability to adjust the imaging parameters is essential to optimize MRI imaging. We show in Figure 1 an example of a pregnancy with TTTS representation in both interfaces. We have recently shown that our platform can help surgeons in their surgical preparations [8] but also for morphological analysis of organs [13]. As the platform is data agnostic, it is also used for visualization and analysis of microscopic data [7]. Finally, we have recently extended the software to allow the analysis of point clouds of any size and origin [4, 5].

3 Segmentating umbilical cords and placenta in twin pregnancies

It is instrumental to have anatomical insight before fetal surgery for TTTS. Hence, the surgeon needs a segmentation of the placenta, the two umbilical cords, the position of the junction point and a fused visualization of the mother, the twins and these segmentations. The surgeon can then evaluate the possibility of the intervention, and define the optimal route to minimize the risks to the fetus and mother, while reaching the critical area.

Manual segmentation of MRI images is both cumbersome and time consuming. The mutual position of the fetuses may partially occlude elements of the cord. In addition, the cord may adopt complex three-dimensional geometrical

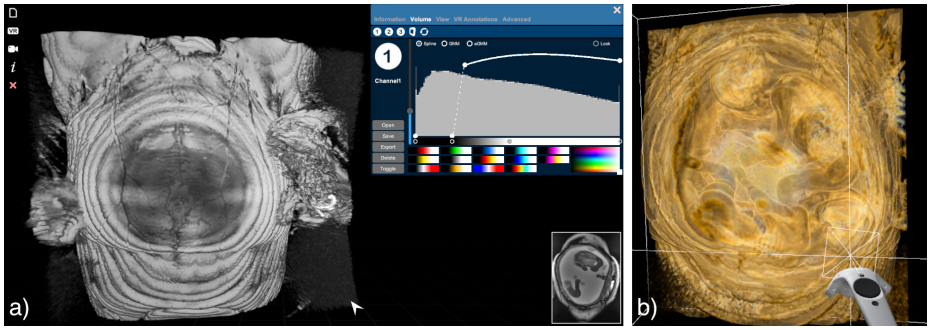


Fig. 1. The DIVA software platform presented through a fetal MRI of monochorial pregnancy with TTTS. a) Desktop (or 2D) interface with raw data in the bottom right corner and its transfer function window in the top right corner. b) VR interface with VR controller in gray.

structures making its annotation complicated. Ideally, this analysis should be done quickly, so that fetuses positions are unchanged since the imaging had been performed.

In order to facilitate the annotation, we have developed a procedure allowing the adaptation of transfer functions to the structures of interest. The visualization in VR is very demanding on the graphics card and the modification of the transfer functions must be done by a computation of limited complexity. We have used a simple kernel-based data representation approach. Through the 2D or VR interface, the user can place a point in his area of interest defining an internal and an external area (see Figure 2). The transfer function is then simply defined as:

$$Tf(v) = \frac{1}{Z}\left(\frac{1}{n}\sum_{i=1}^{n} k(v_i, v) - \frac{1}{m}\sum_{j=1}^{m} k(v_j, v)\right) \tag{1}$$

with n the number of points in the region of interest, m the number of points in the external region, $k(.,.)$ the kernel, v_i the features of the voxels in the region of interest, v_j the features of the voxels outside the region of interest and Z the normalizing constant ensuring that $Tf(v)$ takes value between 0 and 1. In its practical implementation we limited the evaluation to 4 features per voxel: gaussian, gradient, minimum and maximum values over a region of 5 voxels. An example of result is shown in Figure 2. These transfer functions can be modified when changing organs of interest, or within an organ if the images features vary significantly.

Our annotation pipeline is then based on a simplified one-shot learning procedure, designed to minimize the amount of manual tagging of the data. It encompasses several consecutive steps: two-classes tagging in VR, handcrafted features extraction, model training on this subset, and subsequent inference on

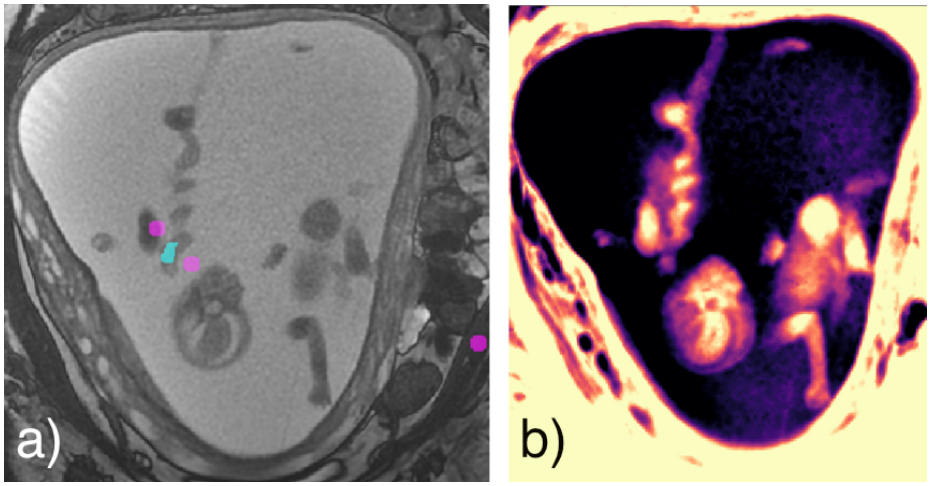


Fig. 2. Improved visualization using transfer function adaptation. a) Coronal section of a raw image. The internal (resp. external) area is indicated in cyan (resp. magenta). b) Resulting image after transfer function transformation.

the whole volume. The procedure, inspired by [2] is described in [9] (Figure 3.a).

VR and optimized visualization parameters are pivotal to ensure fluid tagging. Using the VR controller, the user is able to quickly annotate some voxels belonging to the structure of interest, and background voxels in another color. It is instrumental to tag background voxels belonging to different structures, in order to better refine future segmentation. A clipping plane tool is available to navigate within the data by removing portions of organs in front of the user. A tagging example is presented in supplementary with the clipping plane activated. A limited set of 100 features is associated to each voxel, and the classification procedure is performed on all voxels. This set includes 30 usual spatial filters introduced in [1], as well as 18 random anisotropic filters. Together they capture the close vicinity of each voxel. In order to gain information on objects contours, 12 signed distance fields obtained with diverse image binarizations have been added. We leveraged transfer learning to add global insight on the data. 30 filters are extracted from the penultimate layer of a convolutional autoencoder trained on various raw fetal MRI. The 10 last features are retrieved in the bottleneck of two self-supervised autoencoders [15]. We used two simple pretext tasks consisting on filling missing patches (5-pixel wide) and predicting the vicinity of the voxel patches (5 to 11 pixels wide). We associated the latent space features to the central voxel of the patches and ran the inference on the full stacks. All features evaluations and calculations associated with one-shot learning were performed in the cloud. We used a recently developed library [9] to allow a fluid use of VR while computationally involved calculations were performed on another machine.

4 Results

Image annotations were performed on a Windows 10 based x64 system with an Intel i7-7700 CPU clocked at 3.60 GhZ, with 32 GB of RAM and an NVIDIA GeForce RTX 2080 Ti graphics card. An HTC Vive headset with its controller were used through SteamVR to perform the tagging procedure in VR. Analysis scripts were coded in Python 3.7.

For each of the 14 examples, Dice coefficient was computed between inferred annotation and expert segmentation in order to assess the precision of our procedure. Results of these measurements were 0.59 ($\pm$0.10) and 0.35 ($\pm$0.07) for placenta and umbilical cords segmentation respectively. An example segmentation is available in Figure 3. We provide in Figure 4 a more global view of the fused segmentation of both the placenta and umbilical cords. The position of cords insertion with respect to the mother anatomy is key to the surgery planning. An explanatory video in VR is available in supplementary material.

Those results were obtained after a single round of tagging, training and infer-

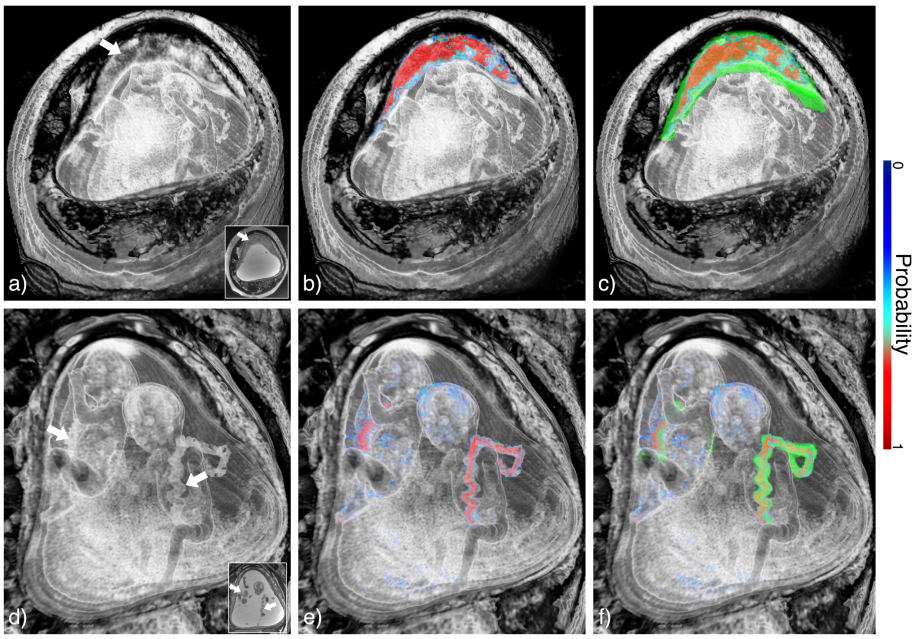


Fig. 3. Segmentation of placenta (top panel) and umbilical cords (bottom panel) on a fetal MRI of monochorial twin pregnancy example. Both structures are indicated with a white arrow. **a-d)** Raw data visualized in 3D and as an image stack in the bottom right corner. **b-e)** Overlay of the raw image in gray and semi-automated segmentation. Colorscale for probabilities is indicated on the right. **c-f)** Overlay of the raw image in gray, output probabilities, and ground truth segmentation performed by an expert.

ence. This whole procedure can however be iterated to correct for suboptimal classification. We performed a second round on a particularly challenging example, increasing Dice coefficient of 0.23 (respectively 0.09) for placenta (respectively umbilical cords) segmentation. Such improvements on an MRI presenting acquisition artifacts prove the robustness of our technique. If the learning procedure saturates, tagging can be finished either manually in VR or corrected from selected 2D slices of the data.

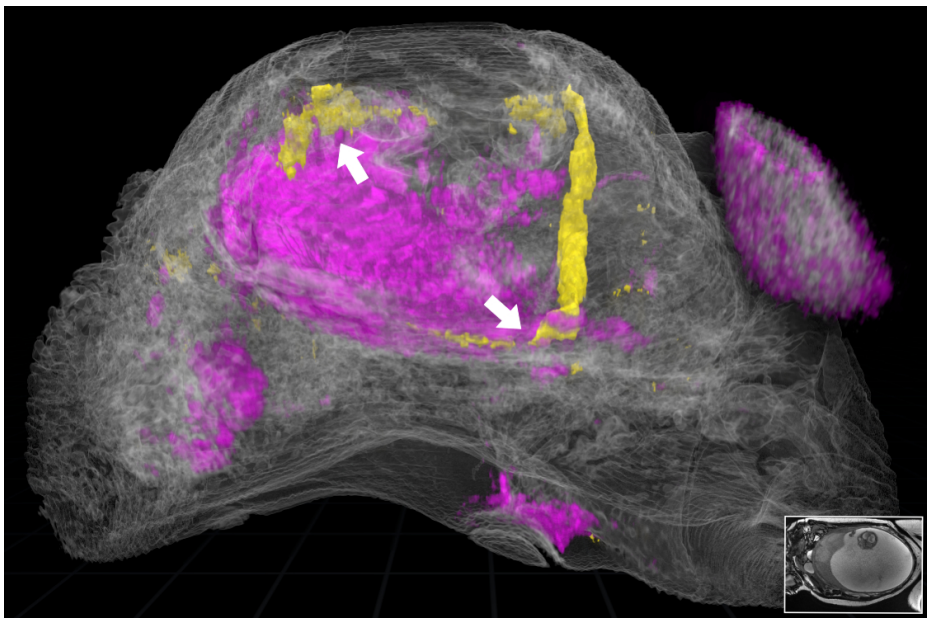


Fig. 4. Fused representation of umbilical cords segmentation (in yellow), placenta segmentation (in magenta) and raw data in gray. A representation as a z-stack is available in the bottom right corner. The insertions of the cords in the placenta are denoted with white arrows.

5 Discussion

Surgeries associated with TTTS are difficult and risky procedures. Optimal surgery for maximum success requires extensive preparation. It is necessary to enter the uterus with the fetoscope in the correct location and orientation to optimize visualization of the chorionic plate, the placenta, and the vascular anasotomoses. This is a prerequisite for complete and selective coagulation of the anastomoses responsible for the syndrome. The ability to plan and simulate surgery, visualize placental location, cords insertions, and anastomoses using modern imaging techniques would be a definite asset in the surgical management

of these conditions. The ability to have a geometric understanding of key organs and to visualize imaging data in a realistic 3D environment will be crucial in defining the way forward with the fetoscope. We have shown that we can characterize and segment the position of key organs using virtual reality and a simple form of machine learning. By overlaying the raw data with the segmented structures, the surgeon can combine multiple pieces of information to ensure that the correct entry point and safest trajectory is found. Future work will focus on automating the detection of geometric constraints to reach the target and provide the possible trajectory of the fetoscope.

References

1. Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., Seung, H.S.: Trainable weka segmentation: a machine learning tool for microscopy pixel classification **33**(15), 2424–2426. https://doi.org/10.1093/bioinformatics/btx180, https://doi.org/10.1093/bioinformatics/btx180, publisher: Oxford University Press (OUP)
2. Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J.I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F.A., Kreshuk, A.: ilastik: interactive machine learning for (bio)image analysis **16**(12), 1226–1232. https://doi.org/10.1038/s41592-019-0582-9, http://www.nature.com/articles/s41592-019-0582-9
3. Bernardo, A.: Virtual reality and simulation in neurosurgical training **106**, 1015–1029. https://doi.org/10.1016/j.wneu.2017.06.140, https://doi.org/10.1016/j.wneu.2017.06.140, publisher: Elsevier BV
4. Blanc, T., El Beheiry, M., Caporal, C., Masson, J.B., Hajj, B.: Genuage: visualize and analyze multidimensional single-molecule point cloud data in virtual reality **17**(11), 1100–1102. https://doi.org/10.1038/s41592-020-0946-1, https://www.nature.com/articles/s41592-020-0946-1, bandiera_abtest: a Cg_type: Nature Research Journals Number: 11 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computational biophysics;Software Subject_term_id: computational-biophysics;software
5. Blanc, T., Verdier, H., Regnier, L., Planchon, G., Guérinot, C., El Beheiry, M., Masson, J.B., Hajj, B.: Towards human in the loop analysis of complex point clouds: Advanced visualizations, quantifications, and communication features in virtual reality **1**, https://www.frontiersin.org/article/10.3389/fbinf.2021.775379
6. Dyer, E., Swartzlander, B.J., Gugliucci, M.R.: Using virtual reality in medical education to teach empathy **106**(4). https://doi.org/10.5195/jmla.2018.518, https://doi.org/10.5195/jmla.2018.518, publisher: University Library System, University of Pittsburgh
7. El Beheiry, M., Godard, C., Caporal, C., Marcon, V., Ostertag, C., Sliti, O., Doutreligne, S., Fournier, S., Hajj, B., Dahan, M., Masson, J.B.: DIVA: Natural navigation inside 3d images using virtual reality **432**(16), 4745–4749. https://doi.org/10.1016/j.jmb.2020.05.026, https://www.sciencedirect.com/science/article/pii/S0022283620303715
8. François, R., Cecilia, N.V., Marie-Paule, V., Jean-Baptiste, M., Mohammed, E.B., Stéphanie, P., Marine, D.T., Laureline, B., Pauline, L., Thomas, B.,

Hossein, K.R.: Management of ischiopagus twin separation with a focus on w–s incision design p. 101747. https://doi.org/10.1016/j.epsc.2020.101747, https://doi.org/10.1016/j.epsc.2020.101747, publisher: Elsevier BV
9. Guérinot, C., Marcon, V., Godard, C., Blanc, T., Verdier, H., Planchon, G., Raimondi, F., Boddaert, N., Alonso, M., Sailor, K., Lledo, P.M., Hajj, B., El Beheiry, M., Masson, J.B.: New approach to accelerated image annotation by leveraging virtual reality and cloud computing **1**, https://www.frontiersin.org/article/10.3389/fbinf.2021.777101
10. Lang, H., Huber, T.: Virtual and augmented reality in liver surgery **271**(1), e8. https://doi.org/10.1097/sla.0000000000003601, https://doi.org/10.1097/sla.0000000000003601, publisher: Ovid Technologies (Wolters Kluwer Health)
11. Pfandler, M., Lazarovici, M., Stefan, P., Wucherer, P., Weigl, M.: Virtual reality-based simulators for spine surgery: a systematic review **17**(9), 1352–1363. https://doi.org/10.1016/j.spinee.2017.05.016, https://doi.org/10.1016/j.spinee.2017.05.016, publisher: Elsevier BV
12. Quero, G., Lapergola, A., Soler, L., Shahbaz, M., Hostettler, A., Collins, T., Marescaux, J., Mutter, D., Diana, M., Pessaux, P.: Virtual and augmented reality in oncologic liver surgery **28**(1), 31–44. https://doi.org/10.1016/j.soc.2018.08.002, https://doi.org/10.1016/j.soc.2018.08.002, publisher: Elsevier BV
13. Raimondi, F., Vida, V., Godard, C., Bertelli, F., Reffo, E., Boddaert, N., El Beheiry, M., Masson, J.: Fast-track virtual reality for cardiac imaging in congenital heart disease **36**(7), 2598–2602. https://doi.org/10.1111/jocs.15508, https://onlinelibrary.wiley.com/doi/10.1111/jocs.15508
14. Salvatore, S.D., Vadalà, G., Oggiano, L., Russo, F., Ambrosio, L., Costici, P.F.: Virtual reality in preoperative planning of adolescent idiopathic scoliosis surgery using google cardboard **18**(1), 199–205. https://doi.org/10.14245/ns.2040574.287, https://doi.org/10.14245/ns.2040574.287, publisher: The Korean Spinal Neurosurgery Society
15. Tomar, D., Bozorgtabar, B., Lortkipanidze, M., Vray, G., Rad, M.S., Thiran, J.P.: Self-supervised generative style transfer for one-shot medical image segmentation http://arxiv.org/abs/2110.02117
16. Tomikawa, M., Hong, J., Shiotani, S., Tokunaga, E., Konishi, K., Ieiri, S., Tanoue, K., Akahoshi, T., Maehara, Y., Hashizume, M.: Real-time 3-dimensional virtual reality navigation system with open MRI for breast-conserving surgery **210**(6), 927–933. https://doi.org/10.1016/j.jamcollsurg.2010.01.032, https://doi.org/10.1016/j.jamcollsurg.2010.01.032, publisher: Elsevier BV
17. Verhey, J.T., Haglin, J.M., Verhey, E.M., Hartigan, D.E.: Virtual, augmented, and mixed reality applications in orthopedic surgery **16**(2). https://doi.org/10.1002/rcs.2067, https://doi.org/10.1002/rcs.2067, publisher: Wiley
18. Walbron, P., Thomazeau, H., Sirveaux, F.: , , virtual reality simulation“ in der orthopädie und unfallchirurgie in frankreich **122**(6), 439–443. https://doi.org/10.1007/s00113-019-0649-1, https://doi.org/10.1007/s00113-019-0649-1, publisher: Springer Science and Business Media LLC
19. Yoo, J.S., Patel, D.S., Hrynewycz, N.M., Brundage, T.S., Singh, K.: The utility of virtual reality and augmented reality in spine surgery **7**, S171–S171. https://doi.org/10.21037/atm.2019.06.38, https://doi.org/10.21037/atm.2019.06.38, publisher: AME Publishing Company