UNIVERSITY OF CALIFORNIA
SANTA CRUZ

FUNCTIONAL MAPPING OF THE MOUSE NEOCORTEX DURING POSTNATAL DEVELOPMENT

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL, AND DEVELOPMENTAL BIOLOGY
with an emphasis in STATISTICS

by

Brian Romney Mullen

June 2021

The Dissertation of Brian Romney Mullen
is approved:

Professor James Ackman, Chair

Professor David Feldheim

Professor Yi Zuo

Quentin Williams
Interim Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

Functional mapping of the mouse neocortex during postnatal development

by

Brian Romney Mullen

Functional imaging of neural cell populations at mesoscale is critical for mapping intra- and inter-regional network dynamics across the dorsal neocortex. Past lab members produced a flexible work flow for producing high quality segmentations of functional structure across neocortex utilizing Independent Component Analysis (ICA). This unsupervised machine learning decomposition of densely sampled recordings of cortical calcium dynamics, results in a collection of components comprised of neuronal signal sources distinct from optical, movement, and vascular artifacts. In this body of work, I built a supervised learning classifier that automatically separates neural activity and artifact components, using a set of extracted spatial and temporal metrics that characterize the respective components. Control data recorded in glial cell reporter and non-fluorescent mouse lines validates human and machine identification of functional component class.

Utilizing the insight from the machine learning analysis, I processed a large dataset from three different transgenic mouse lines expressing calcium indicators in distinct sub-population of neurons: pan-neuronal, upper, and lower layer specific excitatory neuronal subpopulations. I was able to create domain maps to discretize the cortex into functional units and observe how each domain behaves. The application of this data-driven method ultimately reduced the number of time series, while maintaining the informative structure at a mesoscale. From all these

animals, I gain insight into the developing structural changes of circuitry based on age-related functional changes observed in these domain maps.

This work is dedicated to

those curious enough to explore the unknown,

my family and friends,

to my partner, who supports my every endeavor.

Acknowledgments

I would like to thanks James Ackman for his passion and convictions that lead to many discussions of proper science. James was able to push the boundaries of my understandings of the scientific process and was able to get me to be a more adept and practicable scientist. This ultimately prompted me to attain the tools necessary for data exploration shown in this body of work.

I would like to thank Raquel Prado for taking me on as my statistics mentor, allowing me to attain the designated emphasis. Her ability her explain complex statistical theory is extremely admirable, as she has dedicated herself to the education of others and advancement of statistical theory. From the many conversations,I was able to apply the knowledge that I attained and start a collaboration for better models of our data.

I would like to thank Sydney Weiser, without her contributions to the lab, this work would not have been possible. Not only a kind and beautiful human being, but also an adept logical thinker that I look up to. I could not have asked for a better lab mate.

Finally, I would like to thank the two lab manager that came through at my time in the Ackman lab, Sarah Hill and Chad Santo Tomas. Their contributions could not be stressed more. They took on husbandry tasks that freed up my time to dedicate to data collection. Finally, I would like to acknowledge the many undergraduate students that came through the lab that touched my life.

I acknowledge University of California Santa Cruz's Hummingbird Computational Cluster for support and node maintenance and Benjamin Abrams, UCSC life Sciences Mi-

croscopy Center for the use of the Zeiss AxioImager. I also acknowledge the UCSC vivarium staff that assisted in the maintenance of the mouse colonies.

This work was supported by Startup funds from University of California, Santa Cruz, Division of Physical and Biological Sciences, grants from the National Institutes of Health, USA (NIH T32 GM 133391).

The contents of this thesis are currently being prepared for publication.

Chapter 1

Introduction

The development of the neocortex is a dynamic and adaptable process determined by multiple factors. Molecular cues play a pivotal role in proper lamination and regionalization of the neocortex, where transcription factors establish cell identity pushing for further differentiation [23]. As neurons differentiate, their axons grow and are guided via molecular gradients to the proper general regional connections [23, 26]. These two activity independent processes are thought to establish the general cerebral architecture during development [11]. Expanding evidence has shown that neuronal activity is essential for circuit formation and maturation, and that change in activity patterns has direct consequences resulting in neurodevelopmental disorders and learning disabilities [46, 29]. The focus of this body of work is to illustrate the functional activity patterns that participate in the neocortical organization during the first month of rodent postnatal development, which corresponds to the third trimester to adolescence in human development [26, 15].

Two major extra-cortical events that drastically re-shape the functional map of the

neocortex are the innervation by the thalamocortical axons and the integration of cortical GABAergic interneurons migrating from the ganglionic eminences. Thalamocortical neurons extend their axons into the cortex to establish layer 4 targets as early as postnatal day (P) 3 in mice, however these connections continue to refine into the second postnatal week [4, 40, 31]. Thalamic connections are established using molecular guidance, even with mutations in key molecules for synaptic release [41]. In the absence of thalamocortical projections,cortical areas fail to refine, resulting in primary sensory cortices resembling adjacent higher order regions based on genetic profiles [16, 48]. Concurrently, interneuronal populations that originate from the subpallium, tangentially migrate and begin to integrate into the cortex starting around birth [30]. Thalamic inputs continue to refine and assist in neocortical development for the first 12 days by first producing connections to lower cortical layers, extending into layer 4 neurons, and interacting with interneurons to promote connectivity [4, 36, 56].

Large-scale activity patterns that are TTX sensitive emerge during perinatal development [20, 5]. Several notable changes to the activity patterns occur based on developmental progression [70]. At birth, slow but intense cortical early network oscillations (cENO) propagate through the cortex; followed by a faster, more structured cortical giant depolarizing potentials (cGDP) by the middle of the first postnatal week [43, 5]. Restricted early oscillations in functional barrel columns are thought to assist in local circuit development [70, 37, 27, 39]. Of note, evidence shows these functional columnar restrictions occur before the appearance of canonical anatomical representation [69, 33].

Neural activity is required for establishing the microcircuit and long range connections. Functional activity, specifically calcium spiking, helps to establish the microcircuit and

long-range connections through promoting dendrite maturation and axonal outgrowth/branching [14, 72, 68]. Interneuron survival is contingent on the integration with excitatory neurons, sculpting the microcircuit to the needs of the local circuit [65]. When looking at the laminar developmental trajectory, there are distinct cellular properties when comparing physiological recordings from upper and lower layers [52]. It has been shown that layer 4 projects to layers 2/3 early in development, however robust growth and synaptic connections occur between P8 and P16 [8, 10]. Layer 2/3 project to layer 5 establish proper connections by P3 and mature around P13, which resembles the mature circuit [7, 6, 53]. Further, long range intracortical connections require spontaneous and sensory evoked activity for proper connection [38, 59, 54].

Technological advances have allowed a collection of in vivo functional studies have elucidated the regional properties and functional changes across development. Traveling and stationary waves have been reported across all primary regions of the neocortex and are necessary for proper circuit development. Visual systems have co-activating waves across cortex and superior colliculus driven by the retinal activations [24, 17, 2]. Somatosensory and motor regions have co-activating waves across the corresponding body map, coinciding with twitches of the neonate [28, 18]. Standing waves occur in each barrel of the barrel field in the mouse cortex, and circuit formation seemed initially independent of whisker input [39]. Further, several of these reports pointed to the sudden and surprising switch of the cortex responding to its intended stimuli during the second week of development, coinciding with circuit maturation [39, 18].

While primary cortical areas are typical areas of investigations due to access to targeted manipulation of the circuit, mesoscale imaging has shown that all patches of the cortex

are functionally active [3]. This has been accomplished through the advancements in ultra sensitive calcium sensors for recording neural activity and mouse genetic lineages, which allow for recording large number of distinct subpopulations of neurons [13, 25, 35]. Recently, our lab has derived methods to segment these mesoscale imaging session using a data-based approach to explore across the entire the developing cortex regardless of age or genotype [63].

This unsupervised machine learning decomposition of densely sampled recordings of cortical calcium dynamics results in a collection of components comprised of neuronal signal sources distinct from optical, movement, and vascular artifacts. In **Chapter 2**, I explain the supervised learning classifier that we built to automatically separate neural activity and artifact components, using a set of extracted spatial and temporal metrics that characterize the respective components. I demonstrate that the performance of the machine classifier matches human identification of signal components in novel data sets. Further, I analyzed control data recorded in glial cell reporter and non-fluorescent mouse lines that validates human and machine identification of functional component class. This combined workflow of data-driven video decomposition and machine classification of signal sources will aid robust and scalable mapping of complex cerebral dynamics.

In **Chapter 3**, I discuss applications of mesoscale calcium imaging, which enables us to visualize functional circuit maturation in the dorsal postnatal developing neocortex. Utilizing independent component analysis, I create domain maps to discretize the cortex into functional units and observe how each domain behaves. This data-driven method reduces the number of time series that are needed to assess, while maintaining the informative structure at a mesoscale. This method is applied across a large dataset from three different transgenic mouse lines ex-

pressing calcium indicators in distinct sub-population of neurons: pan-neuronal, upper, and lower layer specific excitatory neuronal subpopulations. From all these animals, I gain insight into the developing structural changes of circuitry based on age-related functional changes observed in these domain maps. Individual layer dynamics observed in upper and lower layer populations progressively get faster and have higher frequencies than those seen in the pan-neuronal mouse lines. Further investigation into this rich dataset may reveal more understanding of the developing dynamics in cortical structure.

I finalize this dissertation in **Chapter 4**, where I recap the major findings of this body of work and the reference collaborations and applications that utilized this analysis.

Chapter 2

Automated classification of signal sources in mesoscale calcium imaging

Introduction

A combination of advances in genetic engineering, parallel signal acquisition, and computation have enabled tremendously increased amounts of time series data concerning the function of neural systems [12]. As a result, development of analytic tools to aid in the extraction of meaning from these large datasets is being undertaken. Implementation of unsupervised and supervised machine learning strategies can aid in extraction and interpretation of signals from complex time series data increasingly common in systems neuroscience. Machine learning techniques have been utilized to filter experimental artifacts, perform signal de-mixing, and understand brain-wide functional associations in data recorded with techniques such as fMRI and EEG [51, 58, 22]. While analytic pipelines exist for these datasets, other types of functional

recordings have their own advantages, as well as experimental challenges.

Mesoscale functional imaging across the dorsal cerebral cortical surface is used to map and understand dynamics that bridge inter-regional (macroscale) and inter-cellular (microscale) datasets. Large calcium imaging datasets have been popularized due to advances in mouse transgenics and the ability to record calcium flux activity simultaneously from many neurons [44]. Analyzing population dynamics within and between regional networks in high supracellular resolution recordings can further our understanding of information processing in the brain. Mesoscale calcium imaging has been used to investigate patterns of neural activity that assist in self-organization of the developing brain [2, 3]. Functional imaging in behaving mice has been utilized to identify new cortical subregions [73]. Additionally, wide-field calcium imaging of the neocortex during electrode stimulation has helped map information flow between distant cerebral subregions [66].

Determination of neural signals from calcium imaging sessions in unanesthetized, behaving mice is an essential, but challenging task due to numerous confounding signal sources. Vascular artifacts are commonly seen due to vasodynamics and the resulting changes in blood flow to meet the energy demands of surrounding tissue. Fluid exchange between vascular and neural tissue causes cortical hemodynamics, which results in region specific changes of optical properties among cerebral lobes [34]. Further, during the experimental preparation the skull is typically fixed to a specific location, however slight brain movements occur within the cranium that influence the recordings. Any optical property differences that originate from the experimental preparation may be highlighted in the dataset as signal due to changes in tissue contrast.

Principled exploration and identification of each signal source is necessary to reduce the contamination resulting from these physiological dynamics. Independent component analysis (ICA) is a nonparametric unsupervised machine learning technique that I utilized to identify each signal source in densely sampled (5.5 million pixels per frame) calcium imaging videos based on their spatially co-activating pixels and temporal properties [63]. The global mean time-course was initially subtracted and stored, thereby zeroing out each frame and allowing ICA to decompose each signal distinct from global effects. The decomposition results in hundreds of neural source components per hemisphere that are distinctly de-mixed from from artifact source signals.

In this study, I determined necessary conditions for identifying neural signals in mesoscale cortical calcium flux videos and create a machine learning classifier to isolate neural activity for analysis. We show that a combination of morphological and temporal metrics can characterize the signal components that maximize statistical independence in the video and I built a classifier that automatically selects for neural components. Our concurrent analysis of control wide-field imaging data corroborates the identification of artifact signal sources and gives insight into the structure of neuronal calcium dynamics across neocortex.

Results

ICA was successfully used to extract components from mesoscale imaging data across the dorsal cortex of postnatal day 21 unanesthetized, pan-neuronal GCaMP6s expressing mice. Each component has a spatially decomposed eigenvector with its corresponding relative intensity temporal fluctuations. Similar data was recorded and processed in three sets of age matched control mice: cx3cr1 GFP (microglia; mGFP), adhl1 GFP (astrocyte; aGFP), and the non-transgenic C57/black 6 (Bl6) mice.

Spatiotemporal metrics from components

Inspection of each set of experimental components resulted in the human classification as neural or artifact (**Fig.2.1 a**). The artifacts were further distinguished between vascular and other for descriptive purposes. Neural components typically have globular spatial representation with highly dynamic properties. Vascular artifact components can be easily identified by the vascular-like spatial representation, which are also temporally dynamic. Other artifact components that are commonly seen in the components are movement or preparation artifacts. These typically have a diffuse spatial representation with smaller or sparse temporal activations. We manually scored each component in the dataset as an artifact (vascular or other) or neural component (**Fig.2.1 b**). From all the GCaMP experiments, on average $73.5 \pm 5.9\%$ of the components were identified as neuronal, where the remaining $26.5 \pm 6.3\%$ were artifact (vascular: $8.7 \pm 2.7\%$; other: $17.6 \pm 7.1\%$). GCaMP mice had substantially higher numbers of neural components compared to the controls, resulting in four times as many as the GFP mouse lines and

six times the number in Bl6 mice (mean number of neural components GCaMP: 235, mGFP:62, aGFP:54, Bl6: 39).

We extracted spatial and morphological metrics of the neural and artifact components to characterize spatial feature differences (**Fig.2.1 c**). From each component's spatial eigenvector I can pull general spatial intensity metrics like global minimum and maximum. Further, the largest eigenvector values correspond to the regions that have the most dynamic change after the video rebuild. Given that the shapes of the high pixel intensity values are used by humans to identify their classification, I decided on a dynamic thresholding technique to binarize the eigenvector. When examining the histogram of intensity values of the neural eigenvector (Fig. 1c), there is a large population of pixels centered around zero with a single long tail. We identified all pixels that were unique to the long tail by excluding all values that lie within range of the shorter gaussian tail. From these binarized masks, morphometrics of each primary region of the component can then be quantified, such as the axis lengths or eccentricity of the shape.

We characterized temporal dynamics of each component by extracting features from the corresponding temporal fluctuations and their frequency analysis (**Fig.2.1 d**). The relative intensity fluctuations allows us to pull out temporal features of each component, like standard deviation and global maxima/minima of each component contribution. We performed wavelet analysis on these time series to characterize only highly significant frequencies (**Fig.2.5**). We calculated a power signal to noise ratio (PNR) with the 95% quantile of red noise defined by the autocorrelation value of each time series. With this ratio, significant frequencies resulted in a value above 1.

After extracting these metrics, I then compared the diverse populations of signal and

artifact components, separated between vascular and other, for each feature of interest (**Fig.2.1 e-h**). A full list of all metrics their respective definitions is provided, organizing between spatial, morphometric, temporal, and frequency features (**Fig.2.6**). While the data shows trends, there is not one single metric that alone could predict the classification of artifacts and signal components.

Control neural components are not distinct globular regions as those from the GCaMP line (**Fig.2.7**), but rather had co-activity with vascular units in the center of its domain. This resulted in the thresholded region to be more similar to the vascular artifacts seen in GCaMP components. However, I was still able to find example components that only had vascular spatial representation without the surrounding tissue activation. Finally, I found similar artifacts of the *other* category in the control data that are also present in the GCaMP sets of independent components.

Characterization of spatiotemporal metrics

To investigate how well these metrics captured features of each component, I explored the coverage of the cortical surface with regions identified by the dynamic thresholding technique. Plotting all the contours of one experiment, the major footprint of the component shows the representative space of each component within brain region (**Fig.2.2a**). The majority of defined brain regions are represented by the GCaMP component footprints, with varying amounts of overlap associated with different cortical areas. Control data resulted in sparsely mapped footprints across the cortex. Further, the mapped centroid location of all components that had a thresholded region from all GCaMP experiments shows the neural components have high

densities in the sensory portions of the brain (**Fig.2.2b**).

Thresholded GCaMP neural components have high densities in the olfactory bulbs and posterolateral portions of the cortex, including visual, auditory, and somatosensory systems. There are less dense localization of centroids along the anteromedial portions of the cortex, including motor and retrosplenial cortices. Further, in both the GCaMP and control mice, I see the majority of artifact components localize along anatomical brain vasculature. The major venous systems including the rostral rhinal vein, the superior sagittal sinus, and the transverse sinus show high densities of artifact centroid locations [67]. The cerebral arteries are less consistent in localizing the primary domain of their respective components. We see that many of the other artifacts align with the sagittal and lambda cranial sutures [62].

We investigated the effects of wavelet analysis on feature generation, by sorting each component within each experiment to its temporal standard deviation value. We then ordered each class of components based on variance and displayed a grayscale heatmap of the significant frequencies in each component across experimental conditions (**Fig.2.2c**). Taking the average of each of the global wavelet spectrum across each experiment highlighted the prominent frequencies seen in each classification (**Fig.2.2d**). Prominent GCaMP frequencies are between 0.3 to 3.5 hz, where control dynamics are seen between 0-1hz. Vascular components tend to have the same frequencies of their neural counter parts. Other components typically have faster frequencies (above 3 hz), most likely due to motion artifacts.

We looked at the overall distribution of the class of components across the relative variance position in the sorted index (**Fig.2.2e**). We found significant shifts in distribution in the types of components based on variance. Neural component were found with high variance,

however they significantly tapered off nearing the noise floor. We found the highest percentage of vascular components with high variance, followed by constant low probability throughout the relative variance. The other components had the highest probability close to the noise floor, with low frequency throughout the rest of the relative variance position.

We found that wavelet analysis between all GCaMP experiments showed $95.8 \pm 3.8\%$ of the components had significant frequencies to assess. We found that $98.4 \pm 2.5\%$ of all GCaMP neural components had significant frequencies (**Fig.2.2f**). Vascular artifacts had $93.0 \pm 6.5\%$ and other artifacts had $86.6 \pm 9.4\%$ of components with significant wavelet frequencies within their respective class of components.

Among all GCaMP experiments, $92.3 \pm 5.6\%$ of the components had a thresholded region to assess. However, when I looked at the breakdown of the class of components, I found that $99.8 \pm 0.2\%$ of all neural components had a thresholded region(**Fig.2.2e**). The artifacts had fewer thresholded components, specifically the other classification; vascular artifacts had $96.6 \pm 3.6\%$ and other artifacts had $60.2 \pm 16.2\%$ within their respective class of components had a thresholded footprint. Of note, I found that components with low variance close to the noise threshold had increased probability of not having a domain threshold.

Overall, this indicated that all metrics can be generated for the vast majority of neural signal and vascular artifacts. However, on average, 40% of the other artifacts do not have morphometrics for their components and 13% are lacking significant frequencies. Further, this also shows changes in the relative spatial and temporal variance distributions of each of these components. The neural and vascular components align with known and predictive anatomy and were found primarily in mapped locations of increase variance. The majority other components

have less variance, and therefore less contribution to the original dataset and, of the ones that had a footprint, were typically found along cranial sutures. All control data had fewer footprint and frequency metrics (**Fig.2.2a,c**).

Feature Selection

To build a classifier, I need to identify which metrics could distinguish between neural and artifact components. Correlation of metrics between neural signal and their respective t-statistics identified which features were most likely useful to classify a component (**Fig.2.3a**).
For this process, I randomly selected seven animals from our twelve experiment dataset for determining features for training the classifier. The remaining five were used as the novel dataset for validating the machine learning performance.

We trained the random forest classifier with all metrics to identify the importance of each feature (**Fig.2.3b**). The features having the greatest t-statistic magnitude had the most importance for proper classification. In particular, spatial and morphological metrics were found to have the highest relative importance for component classification. The final list of 10 feature metrics utilized in the machine learning process are shown (**Fig.2.3b**).

Machine Learning Performance

We utilized the common approach of hiding a portion of the data from the learning algorithm to validate efficiency of machine learning and establish hyperparameters (**Fig.2.4a**). Stratified sorting was used to ensure an equal ratio of artifact and neural signal was placed into each subset. We sampled and trained the random forest classifier 1000 times to see the distribution

of results. We found that it performed well by all metrics assessed: mean accuracy of 97.1%, mean precision of 98.4%, and mean recall of 97.6% (**Fig.2.4b**).

After establishing the efficacy of the classifier, I set out to assess the full classifier based on all data points from the machine learning dataset. We projected all features onto the first two components of a singular value decomposition (SVD), mapping both the human classification and the mean classifier confidence for 1000 iterations (**Fig.2.4c**). As expected, I saw distinct neural and artifact clusters in feature space. Interestingly, the two different types of artifacts also separated into distinct portions of the projected feature space. The confidence of the classifier showed very few components between the extremes, illustrated by the top binned confidence value distributions for each human classification (**Fig.2.4e**).

We found $71.2 \pm 0.2\%$ of components were binned in highly confident values for neural signals (left-most bin), and 22.7 ± 0.2 were binned in highly confident values for artifacts (right-most bin) ($7.0 \pm 0.2\%$ for vascular; $15.7 \pm 0.1\%$ for other). This indicates that the classifier exhibits reliable confidence in the decision boundaries.

To assess the efficacy of this classifier, I then tested 1000 iterations of novel data, completely new experiments that were not involved in training the classifier (**Fig.2.4a bottom**). We plotted the resulting 1000 iterations of each experiment separately (**Fig.2.4d**). Notably, I found that the overall results were about the same as the subset classifier: mean accuracy of 96.9%, mean precision of 98.0%, and mean recall of 97.6%. From the histogram of classification frequency, I found similar results to the confidence of the classifier (**Fig.2.4f**). Among all components, 69.7 ± 2.0 were confidently classified by machine learning as neural signal (left-most bin), where $25.8 \pm 1.4\%$ were confidently classified as artifact (right-most bin; $7.5 \pm 0.4\%$

of vascular; $18.2 \pm 1.6\%$ of other). The remaining 5% either resided outside the two extreme bins or were mis-classified. We investigated the locations of the components that consistently showed false positive or false negative (**Fig.2.4g**). The majority of these components were either on the edge of the region of interest for the cortical hemispheres or within the olfactory bulb.

Analysis of artifact signals in global mean

Removal of artifact components will ensure that neural signals are the dominant signal to be analyzed; however, during reconstruction of the data, re-addition of the global mean must occur. Thus I examined the influence that removal of artifact components has on the global mean and how filtration of the global mean should be considered. For example, vascular artifacts associated with the superior sagittal sinus contribute to the global mean and increase the range of signals recorded during periods of motion. Assessment of the global mean from GFP control experiments showed pronounced signal in these slower frequency oscillations, suggesting the use of a highpass filter (**Fig.2.8**). Indeed, I found that application of a high-pass filter with a 0.5 Hz cutoff minimizes these types of global slow oscillations (**Fig.2.9**). This type of filtration should not be applied to each component individually, as there are regional networks reliant on these slower oscillations [57]. Removal of these low frequencies from the global mean gave improved identification of the cortical patch signal sources that contribute to neural activation.

Discussion

We have implemented an automated classification pipeline for selection of the maximally independent neural components in widefield calcium imaging videos of baseline neocortical dynamics. We identified combinations of component features that distinguish neural and artifact signal sources from decompositions of sufficient spatiotemporal sampling and feature extraction in calcium flux videos. Vascular network artifact components were found to exist independently of the globular neural signals (**Fig.2.1**) [63]. Vascular networks exhibited dynamics within similar frequency ranges as neural components, due to the refraction or occlusion of the GCaMP signal by the vasculature. Given that vascular artifacts had overlapping temporal and frequency metrics with the neural components, I found that a set of spatial and morphological feature metrics gave optimal classifier performance during machine assignment of components to different signal classes. Furthermore, I showed another class of movement and optical artifacts that are present in both control and experimental GCaMP recordings. Frequency analysis from these other artifacts had higher power in the faster dynamics (above 3 Hz) and are due to motion. Indeed many of these other artifacts were associated with vascular tissue displacments during animal movements (**Fig.2.10**). Thus, it is therefore necessary to removed these other artifacts during video pre-processing. Similar approaches have been implemented in fMRI studies, where researchers built an ICA-based classifier to scrub motion artifacts [50, 45].

The data rebuild of identified neural components with mean filtration is a statistically valid process for isolating neural signals. Our recordings consisted of large sets of densely sampled image frames having at least 12 bits of dynamic range across pixel intensities. We

hypothesize that these sampling conditions coupled with a strong neuronal GCaMP signal-to-noise ratio optimizes ICA's signal de-mixing ability to isolate functionally discrete patches of cerebral cortex from other physiological signals. In control recordings lacking a calcium sensor to report neuronal dynamics, high quality isolation of signal components was not attained given equivalent video sampling conditions. The dynamics of vascular and neural tissues are energetically and thus physiologically linked and the interplay between the hemodynamic responses and neural signals is known [64, 34]. Even so, I found that neural GCaMP components comprise discrete units across the cortex (**Fig.2.2**) [63]. In tissue expressing a contrast agent, such as GFP, the optical hemodynamics are enhanced and result in widespread regional effects among the cerebral lobes from control animals. Wavelet analysis on the global mean and individual neural components show the dominant signals extracted from GCaMP animals as being in a faster frequency range than cortical hemodynamics ($>$1 Hz). Our results indicate that neural GCaMP signals heavily outweigh the neocortical hemodynamic signals in decomposed independent components of densely sampled wide-field calcium imaging videos.

Functional imaging in unanesthetized, behaving animals gives insight into the nature of physiological processes, however, nontrivial challenges arise during such sessions with intermixed sets of time varying signals. This work demonstrates that signal components having maximal statistical independence captured in sufficiently sampled mono-chromatic calcium flux videos exhibit a combination of spatiotemporal features that allow machine classification of signal type. Implementation of automated machine classifiers for neural signals is practical given densely captured arrays of spatially and temporally variant data gathered from individual subjects.

Figures

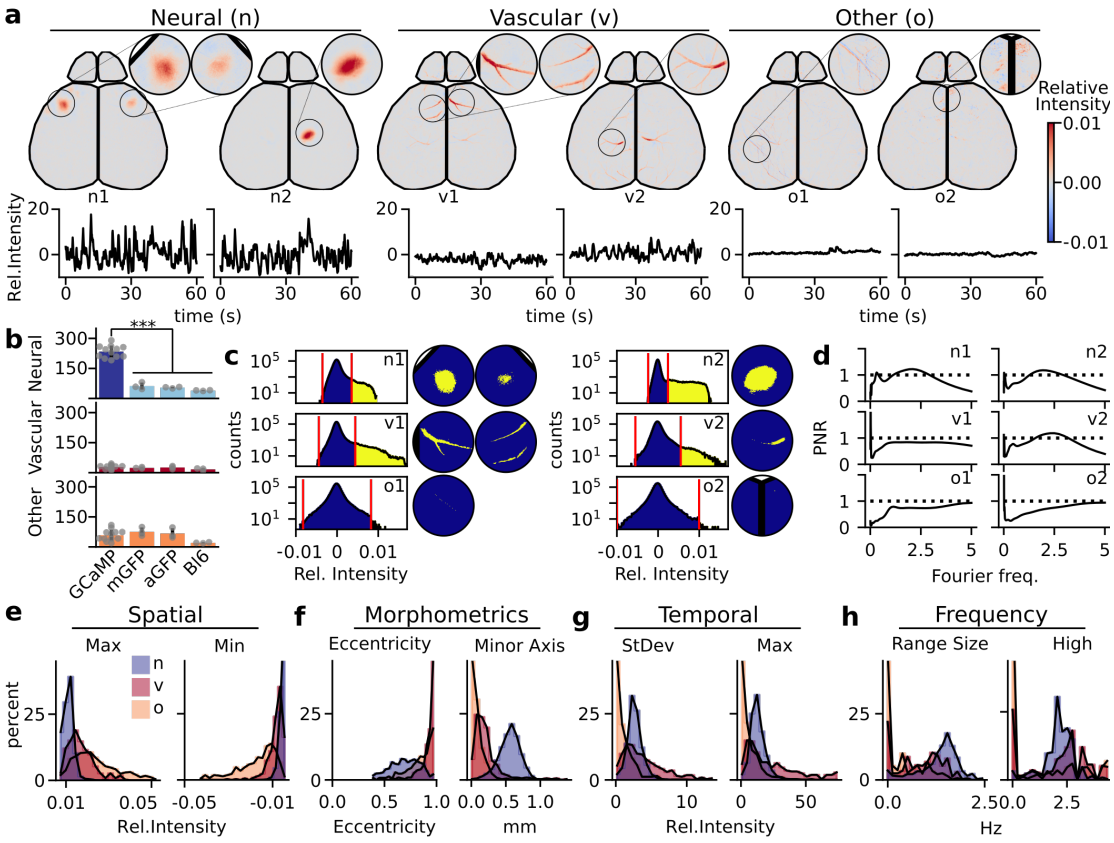

Figure 2.1: Class identity cannot be established by any individual extracted feature. (a) Examples of independent components of neural (n) signal, vascular (v) artifacts, and other (o) artifacts. Components are defined by both the spatial representation (eigenvector) and its temporal fluctuations. Circular windows magnify key portions of the eigenvector. Eigenvector values represented by colormap from blue to red. Temporal representation is in relative intensity (black time course under the eigenvector), only 1 minute of the full 20 minutes are shown. (b) A comparison of the number of neural signal (GCaMP: dark blue; controls: light blue) and the artifact components (vascular: red; other : orange) with each animal shown (GCaMP components: N=12 animals, n=3851; mGFP components: N=3, n = 484; aGFP components: N=3, n = 442; WT components: N=3, n=229). (c) Examples of binarization of the eigenvector. Histogram shows the full distribution of eigenvector values. The dynamic threshold method to generate binarized masks was used to identify the high eigenvector signal pixels (yellow) against the gaussian background (blue). Windowed spatial representation shows binarization on the key portions of the eigenvector. (d) Examples of neural and artifact wavelet analysis shown in the to signal power-noise ratio (PNR) plots. 95% red-noise cutoff was used to create signal to noise ratio (black dashed lines). (e) Histograms of example spatial metrics derived from GCaMP eigenvector values, (f) morphometrics from the shape of the binarized primary region, (g) temporal metrics derived from relative temporal intensities, (h) frequency metrics derived from the PNR.

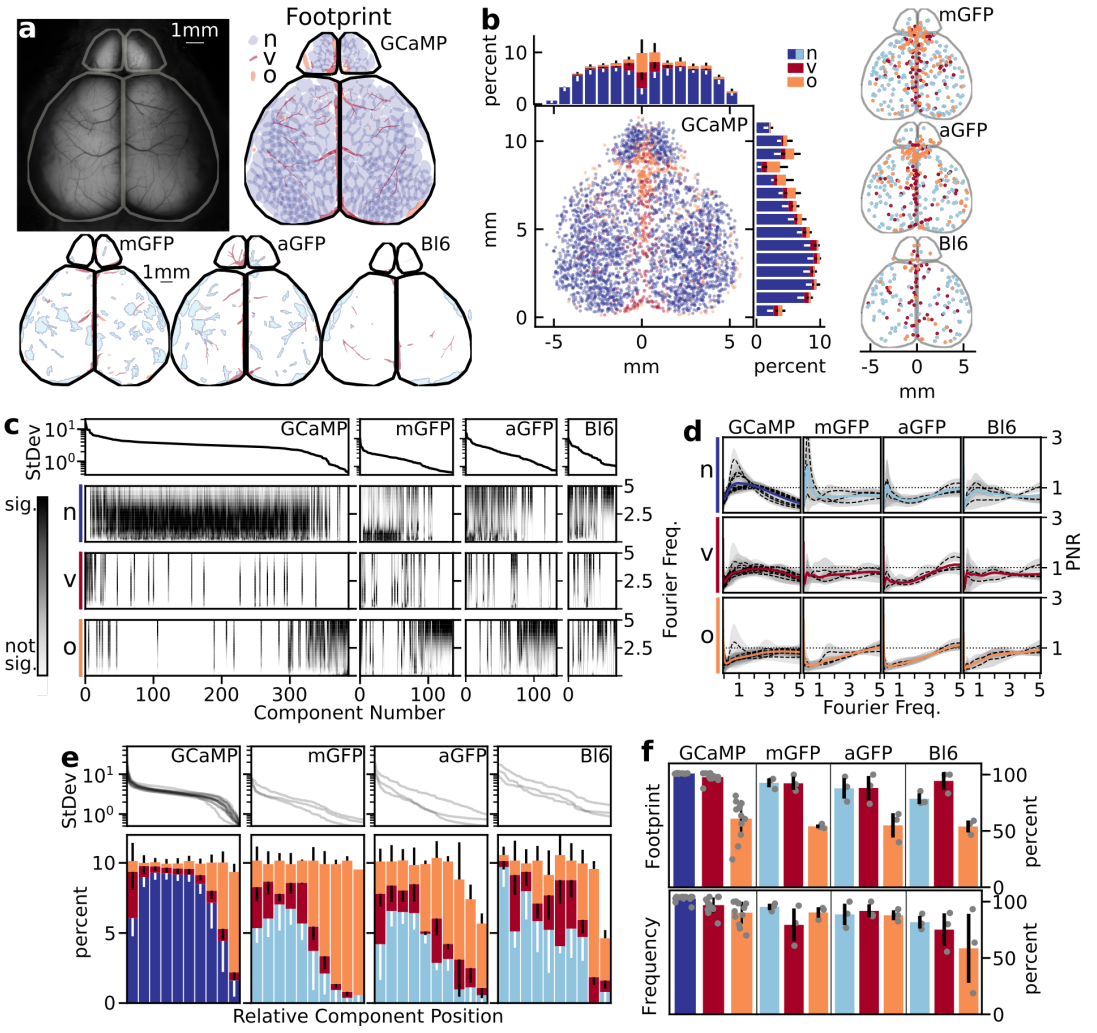

Figure 2.2: Spatial thresholding and frequency data reliably produce neural metrics. (a) Individual experiment preparation with corresponding spatial footprints by class of component: GCaMP neural (dark blue), control neural (light blue), vascular (red), other (orange). (b) All model experiments (N=12) with corresponding centroid location of each of the class of metrics. Histograms show resulting average distribution of spatial location across the field of view (error bars are standard deviation between experiments). (c) Individual experiment (same as (a)), where components are sorted by temporal variance. PNR mapped to each component and organized between the classification of component. (d) Main frequencies seen in each component class between each experimental condition. Dotted lines represent the mean within each animal, where the gray around the mean corresponds to the standard deviation of that animal. The color line corresponds to the grand mean all between experiments. (e) Relative position of the types of components between experiments and transgenic model, shown as the average and standard deviation between experiments. (f) The percent components that had footprints and frequency data that was above the noise cut-off, separated by component type and experimental condition.

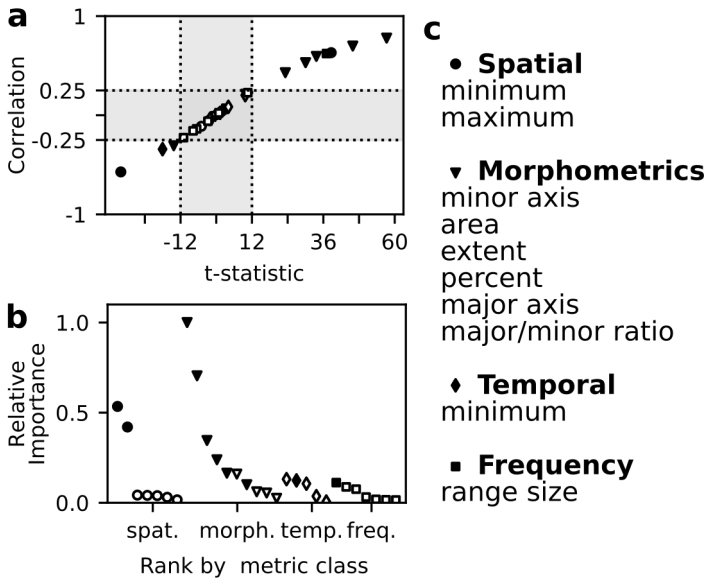

Figure 2.3: Spatial and morphological metrics are most important to classify components. (a) Correlation and t-statistic between artifact and neural components for each feature (N=7, n=2190). Spatial (circles), morphological (triangle), temporal (diamond), and frequency (square) metrics plotted. Cut off values that helped in the selection process are dotted lines, rejected values in gray. Closed points are components that met requirements. (b) Relative importance metric from the Random forest classifier plotted against each metric by their respective classes. (c) Selected metrics shown in the list within each type of feature, sorted by greatest t-statistics magnitude.

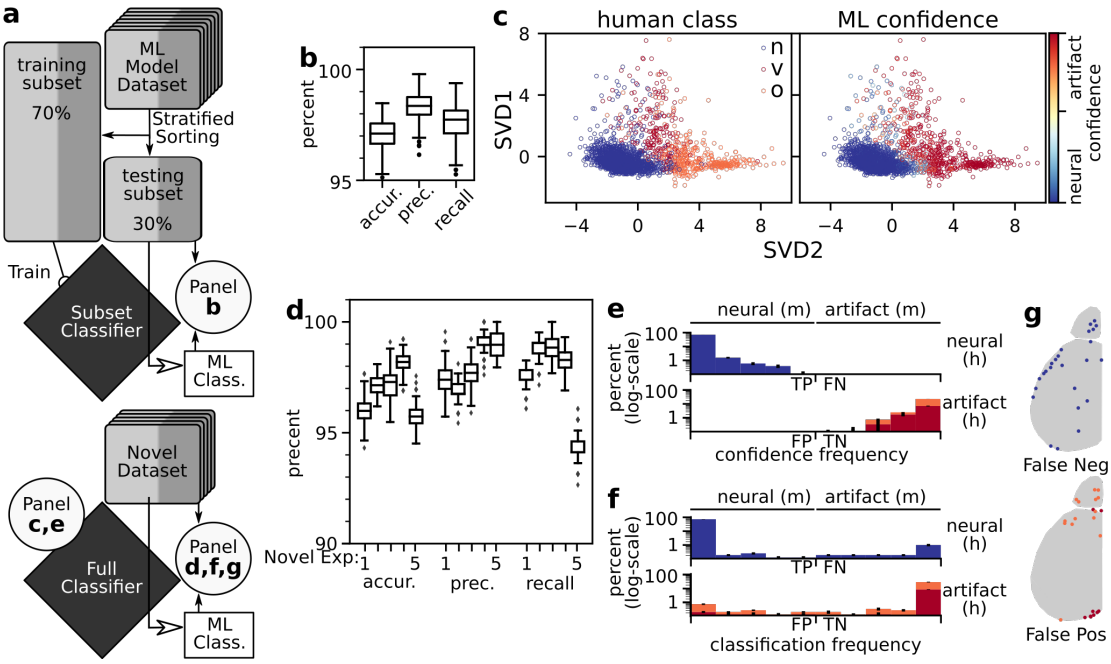

Figure 2.4: Classifier is 97 precent accurate on new experimental data. (a) The dataset was parsed into ML modeling dataset (N=7, n=2190) that was used to establish the machine learning pipeline and a novel dataset (N=5, n=1661) of full experiments that will not influence the classifier. Modeling data was stratified 70/30 split based on classification. 1000 iteration of training the machine learning classifier on selected metrics and validating the machine classification with human classifications. 1000 iterations of training on the full ML building dataset was performed and the novel dataset was assessed on its performance. (b) Performance of the ML training, using subsets of the ML modeling dataset. 1000 iterations resulted in accuracy, precision and recall boxplots. (c) SVD projection of metric data with human classification mapping (left) and the confidence of the ML classifier (right). (d) Performance of each of the novel datasets the full classifier's performance, animals plotted separately showing distribution of the 1000 different trained classifiers. (e) Histogram of human classification with the percent of in each bin confidence bin (same data as c, right). Log-scale was used to highlight the low percentage points. (f) Histogram of human classification with the percent in each binned novel classification. Classification bins based on the percent each classification occurred correctly in the 1000 trials. (g) Approximate location of false negative and positives from novel datasets. True positive (TP), False positive (FP), False negative (FN), True Negative (TN), human (h), machine (m)

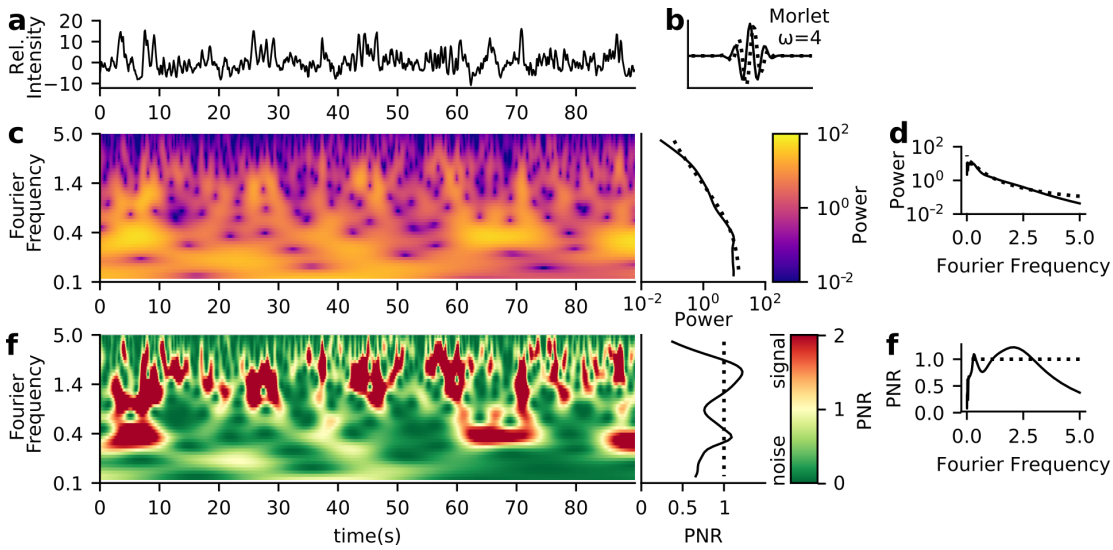

Figure 2.5: Morlet wavelet transform can be used to create a signal to noise ratio that indicates frequencies of high likelihood of signal. (a) Example neural time series, 90 sec of data recorded at 10hz reported in the temporal portion of a component (b) Morlet wavelet (=4) was used for the wavelet transform. (c) The power spectra of the wavelet transform (colorbar, purple to yellow) and the global spectral analysis (black, right). The 95th quantile is shown in dashed lines on the global spectral analysis. Reformatting the frequency spacing, produces (d). (d) A signal to noise power ratio is calculated by dividing the power spectra by the 95% quantile. All values above 1 (dashed line) would indicate a high probability of signal. Anything below 1 would most likely be considered noise (colorbar, green to red). Reformatting the frequency spacing, produces (f).

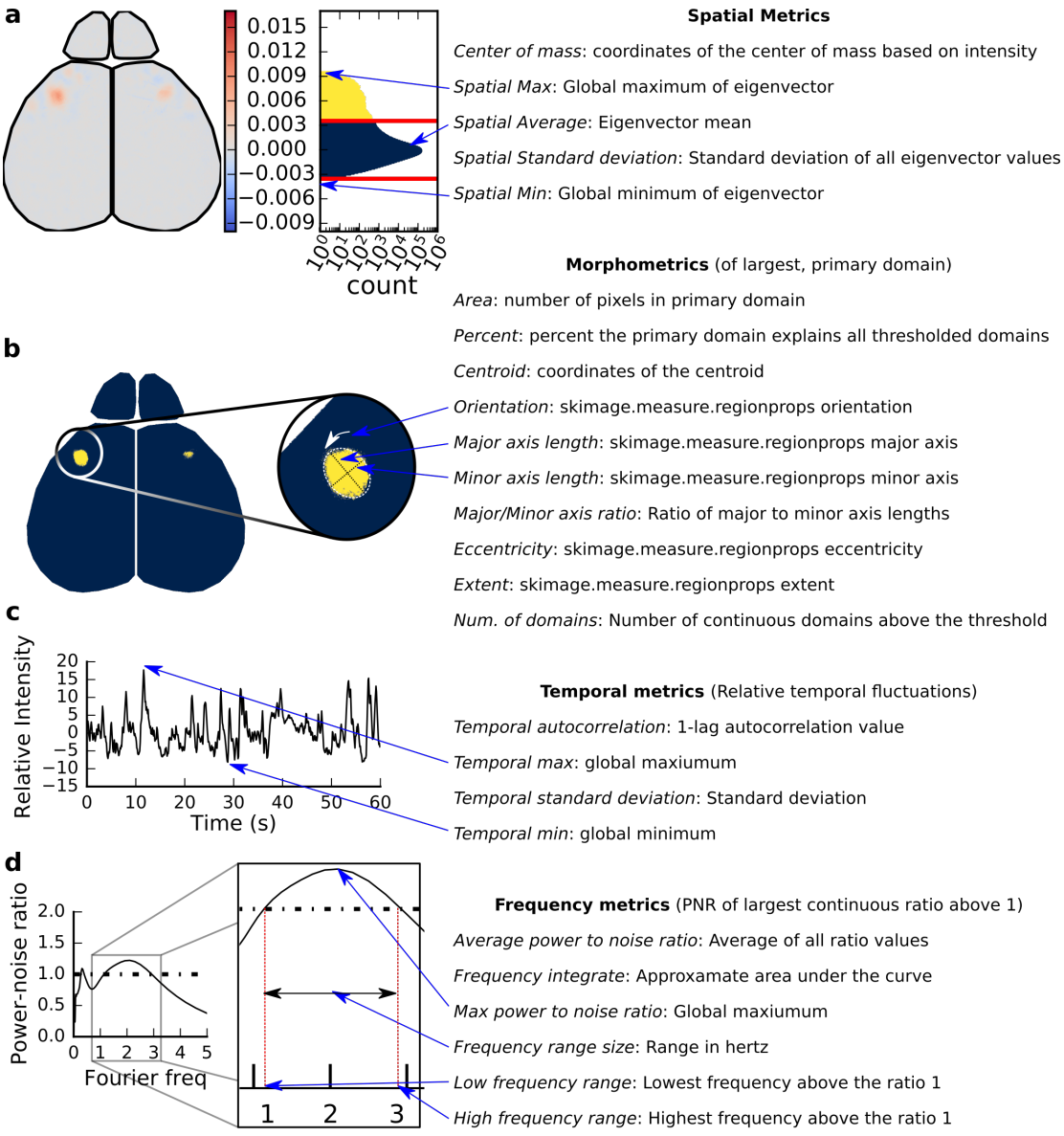

Figure 2.6: Spatial, Morphometric, Temporal, and Frequency features extracted from components. (a) Spatial metrics from statistical characteristics of each eigenvector (spatial representation of the component). The histogram of all eigenvector values is shown the right of the eigenvector. (b) Morphometrics collected from the binarized thresholded masked region of the eigenvector. The largest (primary) domain was used to generate the features for each eigenvector. The majority of metrics calculated utilizes sci-kit image region properties. (c) Temporal metrics are statistical descriptors from the corresponding row of the mixing matrix for each eigenvector. (d) Frequency analysis was done on the mixing matrix row, utilizing the PNR calculated from wavelet analysis (Figure S1). The longest of all continuous frequencies was used to extract each feature.

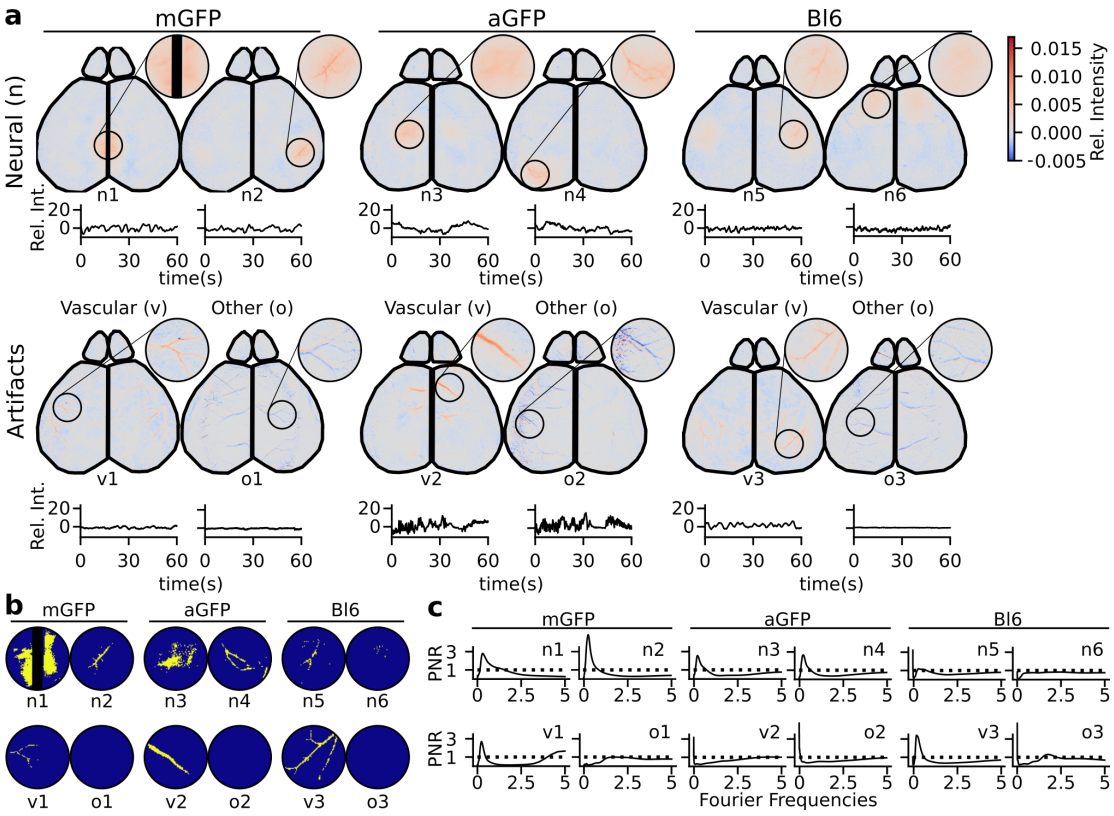

Figure 2.7: Examples of control components, resulting in similar artifact components to GCaMP recordings. Control components from 20 minutes of recording from cx3cr1 GFP (microglia; mGFP, left), adlh1 GFP (astrocyte; aGFP, center), and Black 6 (Non-transgenic; Bl6, right) mice. Two IC examples from each control group corresponding to hemodynamics/neural activity (top) and artifacts (bottom). Artifacts chosen shows a vascular and other artifact commonly seen in GCaMP recordings. Similar data description in regards to temporal and spatial representations as seen in Figure 1. (b) Examples of control binarization of the eigenvector only showing the windowed spatial representation on the key portions of the eigenvector. (d) Examples of neural and artifact wavelet analysis shown in the to signal power-noise ratio (PNR) plots.

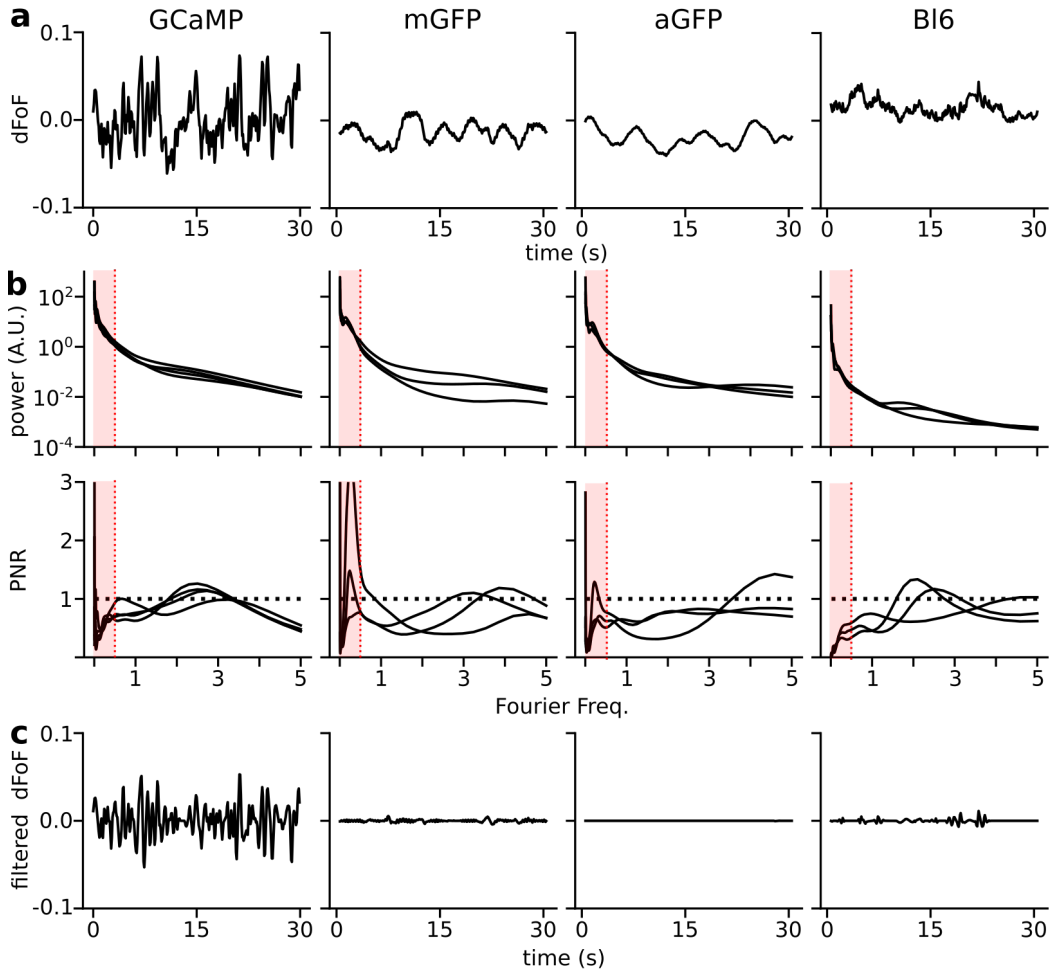

Figure 2.8: Mean filtration to minimize global slow oscillations seen in GFP control data. (a) 30sec examples of the global mean that was subtracted and stored at the initiation of the pipeline, before the decomposition into eigenvectors for GCaMP, mGFP, aGFP and Bl6. (b) Global wavelet spectrum (top) and its corresponding power to noise ratio (PNR; bottom) of GCaMP (N=4), mGFP (N=3), aGFP (N=3), and Bl6 (N=3), red indicates the omitted frequencies from our applied high-pass filter. (c) High-pass filtration results of the same 30 sec in (a).

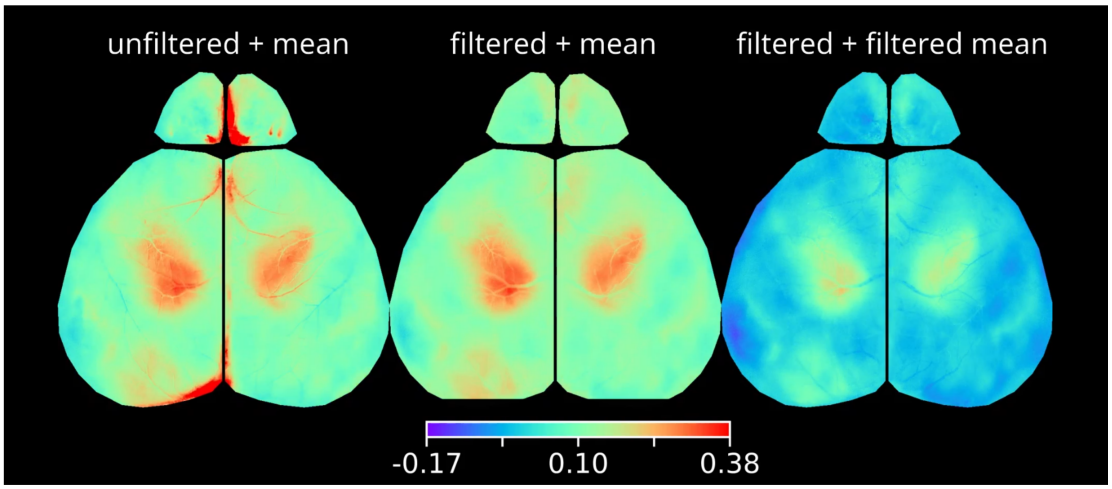

Figure 2.9: Readdition of the global mean with high-pass filter better represents neural signal. Rebuilt video from all components (left) or with only neural components (center, right) with mean re-addition from the original global mean (left, center) or the global mean with 0.5 Hz high pass filtration (right). All movies are on the same scale of change in fluorescence over mean fluorescence (rainbow colorbar)

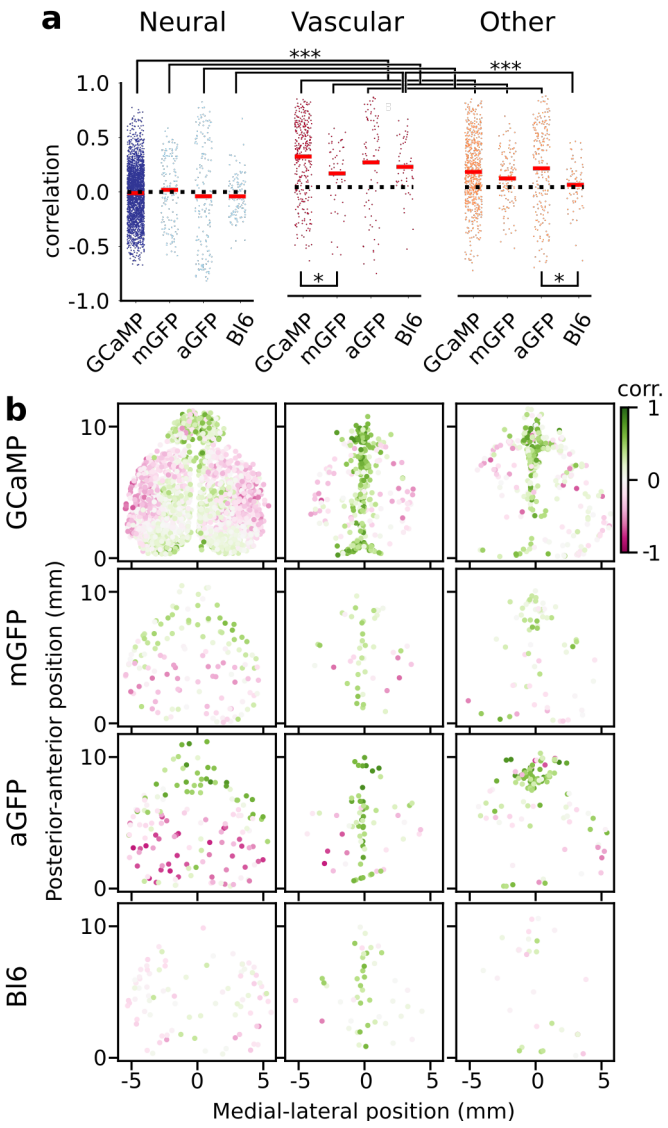

Figure 2.10: Vascular and other artifacts are more correlated to movement than neural components. (a) All neural (blue), vascular (red), and other (orange) components and their correlation to the motion vector from each animal. (b) Spatial location and corresponding correlation (green to pink) of each of component to motion based on their respective classification and genetic background. neural: left column, vascular: center column, other: right column. Top row: GCaMP, second row: mGFP, third row: aGFP, last row: Bl6

Methods

Mice

All animal studies were conducted in accordance with the UCSC Office of Animal Research Oversight and Institutional Animal Care and Use Committee protocols. P21-22 Snap25 GCaMP6s transgenic mice (JAX: 025111), Cx3cr1 GFP (JAX: 005582), and Aldh1 GFP (MGI: 3843271) were maintained on a C57/Bl6 background in UCSCs mouse facilities. To identify Snap25 GCaMP expressing mice, a single common forward primer (5'-CCC AGT TGA GAT TGG AAA GTG-3') was used in conjunction with either transgene specific reverse primer (5'-ACT TCG CAC AGG ATC CAA GA-3'; 230 band size) or control reverse primer (5'-CTG GTT TTG TTG GAA TCA GC-3'; 498 band size). The expression of this transgene resulted in pan-neuronal expression of GCaMP6s throughout the nervous system. To identify GFP expressing mice a forward (5'-CCT ACG GCG TGC AGT GCT TCA GC-3') and reverse (5'-CGG CGA GCT GCA CGC TGC GTC CTC-3'; 400 band size) PCR amplification was used to identify which animals had the GFP transgene. At the end of each recording session, the animal was either euthanized or perfused and the brain dissected.

7 animals used in this study were experiment and control mice a study looking at microbiota influence on cerebral networks.[47] These methods work independent of experimental condition and the perinatal penicillin had little effect on domain parcellation.

Surgical procedure

All mice were anesthetized with isoflurane (2.5% in pure oxygen) for the procedure. Body temperature was maintained at 30C for the duration of the surgery and recovery using a feedback-regulated heading pad. Lidocaine (1%) was applied subcutaneous on the scalp, followed by careful removal of skin above the skull. Opthalmic ointment was used protect the eyes during the surgery. The cranium was attached using to two head bars using cyanoacrylate, one across the occipital bone of the skull and the other on the lateral parietal bone.

After the surgery was complete, mice were transferred to a rotating disk for the duration of the recording. At the end of the recording session, the animal was euthanized and the brain dissected.

Recording calcium dynamics

In-vivo wide-field fluorescence recordings were collected in a minimally invasive manner. Imaging through the skull by single-photon excitation light from two blue LED light (470 nm; Thorlabs M470L3) produces a green fluorescent signal that is collected through coupled 50*mm* Nikon lenses (f5.6 / f1.2, optical magnification $\sim$ 1x) into a scientific CMOS camera (PCO Edge 5.5MP, 6.5μm pixel resolution). The top lens in the tandem lens setup was used to focus on the cortical surface, thereby lowering the magnification slightly; anatomical representation for each pixel corresponded to $6.9 \pm 0.2\mu$m (min: 6.7μm, max: 7.2μm). Excitation light was filtered with a 480/30 nm bandpass (Chroma Technology AT480/30x) and the emission signal was filtered with 520/36 nm bandpass (Edmund Optics 67-044). Data collection was performed

in a dark, quiet room with minimal changes in ambient light or sound, thus the brain activity recorded was resting state without direct stimulation. Raw data was written directly as a set of 16 bit multi-image TIFF files.

The total amount of data recorded for each animal was generally at least 40 min and the amount of time in between video segments was less than 1 minute. All analyzed data consisted of two sequential full spatial resolution recordings concatenated together giving 20 min of data sampled at 10 frames per second for each video decomposition[63].

Dynamic Thresholding

The spatial histogram of eigenvector values across each component can be visualized as a single tailed gaussian distribution centered around 0, where the tail represents the spatial footprint of each component. The two edges of the distribution are first identified. The boundary closer to 0 is taken as the edge of the central noise distribution, and that boundary is used to define the dynamic threshold on the opposite side of 0. Any values outside of this noise distribution and is part of the wider tail are included in the binarized domain of the component. If the tail was negative, the component was flipped spatially and temporally for visualization purposes.

ICA and Data processing

ICA decompositions were run on a single computing cluster node having 1024 GB of RAM. After ICA processing, metric extraction was computed on local computers with 16-32GB of RAM.

Metric generation and classification of Neural Independent Components

An ensemble random forest classifier from the scikit-learn 23 package[1] was used to train and classify between human scored signal and artifact components [21], based on features calculated from each component .

Wavelet Mean filtration

Wavelet decomposition on the time series signals were performed with a $\omega = 4$ morlet wavelet family, code adapted from C. Torrence and G. Compo,[55] available at

URL: http://paos.colorado.edu/research/wavelets/

Significance was determined using the 95th percentile of a red-noise model fit to the time series autocorrelation. Frequency distributions are all displayed as the ratio of the global wavelet specturm, relative to the noise cutoff. For wavelet filtering, the original signal was rebuilt excluding all frequency signals in a certain range.

Statistical significance

Statistical significance was calculated using OLS models from `statsmodel.formula.api` with Holm-Sidak multiple testing correction ($p \leq 0.5$: *; $p \leq 0.01$: **; $p \leq 0.001$: ***). Model significance is determined by the F-statistic, and significance of two-group analyses (p>|t|) are calculated with t-tests.

Chapter 3

Functional Mapping of the Mouse Neocortex During Post-natal Development

Introduction

Neural activity is essential for proper organization and maturation of the cortex, well before behavioral or sensory input. During development, change in activity patterns has direct consequences resulting in learning disabilities and neurodevelopmental disorders [46, 29]. In mice, these spontaneous traveling and stationary waves have been reported across all primary regions of the neocortex. In primary visual cortex, co-activating waves co-exist with the superior colliculus driven by the retinal activations [24, 17, 2]. Barrel cortex has reported standing waves in each barrel in the mouse somatosensory cortex, and circuit formation seemed initially independent of whisker input [39]. Somatosensory and motor regions have waves across the corresponding body map, coinciding with twitches of the neonate [28, 18]. Further, several of

these reports pointed to the sudden and surprising switch of the cortex responding to its intended stimuli during the second week of development, coinciding with circuit maturation [39, 18].

While primary cortical areas are typical areas of investigations due to access for targeted circuit manipulation, mesoscale imaging has shown that all patches of the cortex are functionally active as early as P3-5 [3]. This has been accomplished through the advancements in ultra sensitive calcium sensors for recording neural activity and mouse genetic tools, which allow for recording large numbers of distinct subpopulations of neurons [13, 25, 35]. Recently, our lab has developed methods to segment these mesoscale imaging sessions using a data-based approach to explore across the entire developing cortex regardless of age [63]. This method discretized the cortical sheet into functional units based on recorded activities, resulting in a domain map of a few hundred domains. These domains can be assessed based on function and connectivity.

In this study, I have established development functional profiles for three targeted neuronal populations using upper-layer (cux2), lower-layer (rbp4), and pan neuronal (snap25) driven expression of the genetically encoded calcium indicator (GCaMP6). I provide a glimpse into the functional dynamics of the entire dorsal cortex through the first three weeks of development. I describe functional profiles for both primary cortical regions as well as associated cortices, in attempts to understand the development of higher order functional networks. Further, I highlight varying types of information based on the subpopulation of neurons utilized. Finally, these profiles can be used to approximate functional mapping for experiments, or will assist researchers with methods to determine the functional map based on their own experimental conditions.

Results

Transgenic expression profiles of calcium indicators

To explore the developmental progress of functional activity, I chose the *snap25* $^{+/snap25\text{-}2A\text{-}GCaMP6s}$ (snap25 GCaMP6s) and recorded functional activity throughout the first three weeks of development. This was done to observe the pan-neuronal progression activity in cells capable of synaptic release [60]. Previous work did not characterize if the snap25 protein is expressed in the transgenic mice , given the self proteolytic capabilities of the gene construct leaves protein residue on snap25 protein (**Fig.3.7 a**). [35]. I investigated the developmental protein expression profile and showed dynamic changes for all three proteins: snap25, GFP, and GCaMP. All protein products follow similar developmental profiles, increasing in the first 2 weeks and stabilizing after the second week, consistent with known profiles (**Fig.3.7 b-c**) [49]. Immunostaining for GCaMP6s and its resulting GFP degradation shows almost uniform expression across the cortical layers, however the younger mice tend to have biased lower layer expression (**Fig.3.7 d-e**). Previous reports on developmental mRNA expression across cortical lamination showed similar results [9].

I further compare between the stated pan-neuronal mouse line (*snap25 GCaMP6s*) to layer specific neuronal lines. Two cre lines were chosen, one for upper layers (layers 2-4; *Cux2-creER*) and the other for lower layer (layer 5; *Rbp4-cre*). Both cre lines were independently mated and back-crossed with *R26 floxed GCaMP6f* $^{+/+}$ (*Ai95*) producing homozygous expression of GCaMP6f in their respective cortical layers. The activities of *Cux2-CreER* and

Rbp4-Cre have been reported in adult mice [25], however their activities in neonatal and young mice are less known. Immunostaining showed of *Rbp4-cre* directed GCaMP6s specifically in deep-layer neurons as early as P6 (**Fig.3.8 e-f**). After tamoxifen induction, GCaMP6s expression was induced in upper-layer neurons at postnatal day (P) 6 and P14 in mice containing *Cux2-CreER* (**Fig.3.8 a-d**). I was able to achieve robust lateral expression at P6, but lacked many of the medial expression; however, by P14 there was an increase in medial expression in addition to the robust lateral expression of the GCaMP6.

Domain map creation and characterization

To attain maximal spatial decomposition of our recorded signal, I verified that I collected enough data at each age [63]. The decomposition results in components made of the spatial representative eigenvector and their corresponding temporal influence on the data [42]. The maximal decomposition was validated by comparing the number of components achieved from 20 minutes to 30 minutes of recording (**Fig.3.9 d-e**). I observed that the recordings that had more components at 20 minutes, increased slightly at 30 minutes. Interestingly, the two extremes of our two week dataset (P2 and P14), did not benefit from more data, underscoring that the temporal complexity of the neural signal (see **Fig.3.3**) does not influence the number of components. Spatial complexity of early functional activity increases due to increasing cell populations expressing GCaMP, however, each cortical patch seems to behave independent of each other. As the animal ages and spatial complexity changes due to the cortex establishing the macro-circuit, it appears as though less data would be required to maximally segment across the cortex. As such, I decided to run the analysis on 30 minutes of data for every age of the first

two weeks of development.

Comparing the eigenvectors of the decompositions, I observed many similarities between eigenvector patterns of different ages and cellular subpopulations. From the data collected at the various ages of the *snap25 GCaMP6s* mice, I identified seemingly corresponding eigenvectors in purported motor (top row), somatosensory (middle row) and higher order visual cortex (bottom row) (**Fig.3.1 a**). Intriguingly, I found similar eigenvectors in recordings of both upper-layer (*Cux2-CreER; Ai95*) and lower-layer (*Rbp4-Cre; Ai95*) dataset (**Fig.3.9 a-c**). Unfortunately, these types of comparisons are most easily done on eigenvectors with multiple areal influences which are a minority of eigenvectors. I found that on average 65-85% of the eigenvectors have only one eigenvector area of high influence and this percentage changed with age (**Fig.3.9 f**); P10-P12 had the most multi-areal eigenvectors (35%), where P2-6 had the fewest (15%).

To ensure that these methods are truly identifying functionally relevant structure, I recorded from 3 different ages (P7, P14, P21) of two mice GFP-expressed uniformly across the cortex, microglia (mGFP; *cx3cr1 GFP*) and astrocytes (aGFP; *aldh1 GFP*). Further, I recorded non-transgenic background line, C57/Bl6 (bl6) mice at those same ages. I processed the video data in similar fashion to the GCaMP6 videos and compared the resulting components and domain maps to the age-matched experimental data.

I maximally project through all eigenvectors to create domain maps that segment the cortex based on the underlying data [63]. Example domain maps from various ages of pan-neuronal (snap25), upper cortical layers (cux2), lower cortical layers (rbp4), and one control (aGFP) are shown (Figure 1 b). I found that all age matched, non-GCaMP controls had non-

significant similar domain numbers (N = 24,two-way ANOVA on age and genotype, age-F:3.39 $p > 0.05$, genotype-F:1.58 $p > 0.05$, interaction-F:0.34 $p > 0.05$), as such I compared all age-matched animals to the collection of controls. Plotting all *snap25 GCaMP6s* animals (N=42) number of domains with respect to age, there is an increase in the number of domains from P2-P5, the domain number plateaus between P7 to P12, then decreases in number to P21 (Figure 1 c-left). However when I ran the ANOVA I found no significant differences between ages based on domain number (one-way ANOVA, age-F:2.85 $p > 0.05$). I compared the data recorded from control mice to those from *snap25 GCamP6s* mice the domain numbers of all age-matched controls were significantly lower than all *snap25 GCamP6s* mice (two-way ANOVA on age and GCaMP6s expression, age-F:10.8 $p < 0.0001$, expression-F: 593.93, $p < 0.0001$, interaction-F:50.66 $p < 0.0001$).

Given that during development the cortical area increases with respect to age (**Fig.3.10**), I investigated how domain density was impacted. I summed the total area for all domains that represent cortical area, and used that to calculate the density of domains (**Fig.3.1 d-left panel**). Again, I found that all age matched, non-GCaMP controls had similar domain densities with each other and were significantly reduced from all GCaMP recordings (two-way ANOVA on age and genotype, age-F:2.18 $p > 0.05$, genotype-F:6.544 $p < 0.05$, interaction-F:0.99 $p > 0.05$; post-hoc on age produced no significant differences). Plotting all *snap25 GCaMP6s* animals (N=42) number of domains with respect to age, there were significant differences between ages (one-way ANOVA, age-F:98.72 $p < 0.0001$). I initially observed an increase in domain density from P2-P4, which peaked between P4-P6. Following this peak, the domain density rapidly decreased until P10, when the change in density of domains slowly decreased between

P10-P12. There was a rapid decrease in density between P12 to P14, and a subtle difference in density between P14 and P21 animals. The domain densities in all age-matched controls were significantly lower when compared to *snap25 GCaMP6s* expressing mice (two-way ANOVA on age and GCaMP6s expression, age-F:33.73 $p < 0.0001$, expression-F: 605.32 $p < 0.0001$, interaction-F: 109.19 $p < 0.0001$).

When I compared the data obtained from the *snap25 GCaMP6s* to those from the *Cux2-CreER; Ai95* (upper-layer neurons) and from the *Rbp4-Cre; Ai95* (lower-layer neurons), I found domain numbers and densities were similar at P6. However, maps in *Rbp4-Cre; Ai95* mice reduced in both domain number and density at older ages (**Fig.3.1 c-d**, right panels). (Domain number two-way ANOVA on age and GCaMP6s driver, age-F:94.98 $p < 0.0001$, driver-F: 20.86 $p < 0.0001$, interaction-F: 2.06 $p > 0.05$; Domain density two-way ANOVA on age and GCaMP6s driver, age-F:94.97 $p < 0.0001$, driver-F: 20.85 $p < 0.0001$, interaction-F: 2.06 $p > 0.05$). Similar to the *Snap25-GCaMP6s* mice, there were significant and substantial decreases between P6 and P10 domain number and density for *Rbp4-Cre; Ai95* mice. *Cux2-CreER; Ai95* mice started off with high density of domains at early ages, especially in the lateral regions where GCaMP6f expression was high, and then became comparable to the *Snap25 GCaMP6s* animals by P14. Significant decreases in density occurred at between both age steps in the *Cux2-CreER; Ai95* mice, where there was only a significant decrease in the domain count between P10 and P14. To determine the effect of TAM on cortical domain maps, I administered TAM to early neonatal *snap25 GCaMP6s* mice through oral gavage, and recorded functional activity at P6. I did not see any significant changes in domain number or density in TAM dosed *snap25 GCaMP6s* animals treated with TAM when compared to the non-treated animals.

Next I examined the distributions of domain sizes and shapes per domain map. For full accounting of the inter-animal variation, I plotted the mean and 95% confidence interval of the full distribution across the domain map for each animal. I found that the domain size for our pan-neuronal line gradually increased on average after P6, with significant one day growths beginning on days 6, 9, 12, and 13 (**Fig.3.2 a,c**, left panel) (one-way ANOVA, age-F:688.89 $p < 0.0001$). In addition, there were significant increases in area from P14 to P21. All non-GCaMP controls showed drastically and significantly larger domains than age-matched GCaMP-expressing mice (two-way ANOVA on age and genotype, area-F:20.79 $p < 0.0001$, genotype-F:906.29 p=0, interaction-F:12.69 $p < 0.0001$).

Similar significant increases in domain areal sizes were observed in layer specific GCaMP6s-expressing mice (two-way ANOVA on age and GCaMP6s driver, age-F:2362.12 p=0, driver-F:677.11 p=0, interaction-F:206.27 p=0). However, a distinct sharp increase in domain sizes was observed from P6 to P10 from Rbp4-Cre; Ai95 mice, whereas Cux2-CreER; Ai95 mice showed a sharp increase in domain size from P10 to P14. Domain sizes increased at a slower pace in Snap25-GCaMP6s mice than either of these drastic changes, but the rate of increase was more similar to the Rbp4-Cre; Ai95 mice. This is illustrated by comparing the significant areal change and the changes to major/minor axis of the pan-neuronal (black), lower layer (orange), and upper layer (blue) signal (**Fig.3.2 a**, right panel). As a control, I found that TAM dosed *snap25 GCaMP6s* did not show significant changes in domain areal sizes from the non-dosed counterpart.

In addition to the size, shape is a key factor of each domain where I show a marked increase in eccentricity at the end of the first week of development (**Fig.3.2 b**). Eccentricity

of each domain measures the severity of deviation away from being circular to elongated (0: circular, 1:highly elongated). The Snap25-GCaMP6s mice showed similar shape distributions at P7, but a noted shift in the average value was noted in the mice older than P8 (one-way ANOVA, age-F:28.74 $p < 0.0001$). This same phenomenon was observed in the Cux2-CreER; Ai95 and Rbp4-Cre; Ai95 mice, and there was an apparent the shift upward in eccentricity from P6 to P10 or P14 (two-way ANOVA age and GCaMP driver, age-F:69.96 $p < 0.0001$, driver-F:5.25 $p < 0.01$, interaction-F: 3.25 $p < 0.05$). There were no significant differences in eccentricity in any mouse line from P10 to P14, nor significant differences between TAM dosed and the non-dosed Snap25-GCaMP6s mice.

I examined spatial locations of the 25th percentile of the largest domains (left) and the 25th percentile most eccentric domains with their respective orientation (right), and found that they revealed underlying cortical structures (**Fig.3.2 d**). While some of the largest and most eccentric resided along known physical borders (ie. edge and midline of the cortex), many others corresponded to known structural regional distinctions. Large domains populated the anterior and medial aspects of the cortex, where eccentric domains were found pushing up against known cortical domains. Compared to the Allen Brain institute reference map for adults, there was a striking correlation between these two factors and the reference map. Further, I examined how these two features developed with respect to age and subpopulation (**Fig.3.2 e**). Conservatively, these distinct boundaries were observed as early as P6; with instances of structure occurring in even earlier data.

Wavelet transform

Utilizing domain maps, I was able to reduce the number of time courses from millions of individual pixels to hundreds of domains, while maintaining a data-driven understanding of underlying cortical structure. This produced a large number of time courses that need to be analyzed based on varying dynamics between each age. Examples of the time courses across developmental ages of *snap25 GCaMP6s*, across a single hemisphere and sorted from anterior to posterior were shown in **Fig.3.3 a**. I observed distinct changes in the amount of activations and the propagating waves across regional areas.

To explore functional activity of each cortical domain, I implemented wavelet analysis to determine prominent frequencies and when each domain was active. Examples of time courses were chosen based on the purported primary somatosensory cortex (top row), retrosplenial cortex (middle row), and primary visual cortex (bottom row) (**Fig.3.3 b**). Their respective domain number is shown in the upper right corner of the time course in **Fig.3.3 a**. I utilized a Morlet wavelet ($\omega = 4$) for this analysis and showed the resulting wavelet transform ratio for each representative time course. To create this ratio, I approximated the distribution of red noise with the same lag-1 autocorrelation value as the time course being analyzed. Dividing the wavelet transformation power by 95 % quantile of the power distribution of this red-noise approximation, I expected any value above 1 would most likely be part of the desired signal.

From the wavelet transform, I projected through time to determine prominent frequencies of each time course, obtaining the global wavelet power spectrum (GWPS) ratio (**Fig.3.3 b**, right of each wavelet transform). The higher the value above 1 (dashed line) indicates the more

likely the frequencies exist in the time course. This allows us to quickly assess the frequencies present in each dataset (**Fig.3.4**; next subsection).

In addition, I projected through frequency space (0.1 - 5 Hz projection) to determine the scaled average wavelet power (SAWP) and its corresponding threshold of significance (dashed line; **Fig.3.3 b**, bottom of each wavelet transform). Any SAWP value above this significance line showed periods of time when the domains were more likely to be active. This was used to characterize the events that occur in each domain (**Fig.3.5**; subsequent subsection after next).

Prominent frequencies across development

GWPS ratios (**Fig.3.3 b**, right of each wavelet transform) revealed that at all ages of the first two weeks of development, the wavelets in *snap25 GCaMP6s* mice were dominated by slower frequencies, typically around 1Hz and lower (**Fig.3.4 b** a). This was observed by highlighting the significant frequencies above the waveform for any average power to noise ratio (PNR) above the cut-off value (1; red bars) and the frequencies corresponding to the local maxima of each individual GWPS ratio (histogram, right column). From the histogram of local maxima, I found that there were typically two distinct peaks that fell above and below 0.75 hz. Further, I observed a transient shift to increased frequency power from P7-P10, which subsequently returned to the prominent slower frequencies at the end of the two weeks of development (P12-P14).

I observe this transient upward frequency shift around P10 in all mouse lines: *snap25 GCaMP6s* (black), *Cux2-CreER; Ai95* (blue), and *rbp4-Cre; Ai95* (orange) mice. The *Cux2-*

CreER; Ai95 and *Rbp4-Cre; Ai95* mice showed the similar two peak distributions at early ages (P6), although the higher peak in *Cux2-CreER; Ai95* animals was shifted to faster frequencies (**Fig.3.4 b**). The frequency shift was detected in both *Cux2-CreER; Ai95* and *Rbp4-Cre; Ai95* mice (**Fig.3.4 b**), that subsequently returned back to similar frequency as in the younger animals (**Fig.3.4 a,d**). There was no difference due to the TAM dosing (green) when compared to the *Snap25 GCaMP6s* animal that did not receive the gavaged drug.

It was not until P21 *Snap25 GCaMP6s* that I observed a significant increase in higher frequency and a decrease in lower frequency dynamics than younger animals, as seen from the GWPS and histograms, as well as the integration of the power spectra from the defined windows (**Fig.3.4 a,e**). Window ranges were selected based on the local maxima histograms, where breaks between peaks resided (0-0.1, 0.1-0.75, 0.75-3, 3-5 Hz). When compared to the younger mice, P21 mice showed significantly lower integrated frequency power in the lowest frequency range (0-0.1 Hz) than any of the younger animals (one-way ANOVA age-F:663.58 p=0). There was an increase in power in the second frequency window (0.1-0.75 Hz) from the first week (P2-P6) into the second (P7-P10), which then shifted lower again (one-way ANOVA age-F:133.53, p=0). Where in the two faster frequency windows (0.75-3 Hz and 3-5 Hz), the integrated power was substantially higher in the P21 animals (0.75-3 Hz one-way ANOVA age-F:2415.09 p=0, 3-5 Hz one-way ANOVA age-F:4941.68 p=0).

Cux2-CreER; Ai95 mice showed an increasing amount of faster frequencies in neural activity compared to the *Snap25-GCaMP6s* and *Rbp4-Cre; Ai95* mice. Slower frequencies in the *Cux2-CreER; Ai95* and *Rbp4-Cre; Ai95* mice tended to be lower than in the *Snap25 GCaMP6s* animals, however, progressive increase from P6 to P10 and then decrease from

P10 to P14 in power frequency were noted in all animals (0 - 0.1 Hz two-way ANOVA on age and GCaMP driver, age-F:1831.78 p=0, driver-F:1748.91 p=0, interaction-F:390.33 $p < 0.0001$; 0.1 - 0.75 Hz two-way ANOVA on age and GCaMP driver, age-F:533.00 $p < 0.0001$, driver-F:5100.93 p=0, interaction-F:94.64 $p < 0.0001$). While *Rbp4-Cre; Ai95* mice had faster frequencies identified by the local maxima of the GWPS ratio than *Snap25 GCaMP6s* mice, the integration showed that neural activity in *Cux2-CreER; Ai95* mice increased with age and surpassed in power in the widowed faster dynamics (both 0.75-3 and 3-5Hz) than the other two mouse lines (0.75-3 Hz two-way ANOVA on age and GCaMP driver, age-F:1424.08 p=0, driver-F:5581.48 p=0, interaction-F:334.667 $p < 0.0001$; 3-5Hz two-way ANOVA on age and GCaMP driver, age-F:2176.17 p=0, driver-F:6148.84 p=0, interaction-F:349.59 $p < 0.0001$). Both *Snap25-GcaMP6s* and *Rbp4-Cre; Ai95* mice showed a slight but significant increase from P6 - P10, however the change was not significant going up to P14. In all frequencies, there was no significant difference due to TAM dosing of *Snap25 GCaMP6s* mice.

Wavelet derived event characterization

SAWP was utilized to determine each activation event, allowing us to characterize all events and showed the progressive increase in activity in each domain of the cortex. To minimize the influence of neighboring domains, the center 100 pixels (10x10 pixels) of each domain was utilized to create the event-based characterization. Each event was defined by SAWP between the 0.1-5 Hz frequencies with powers that resided above the significance threshold of each wavelet transform (**Fig.3.3 b**, bottom of each wavelet transform). This binarized version of the time course, which allowed us to calculate the number of events that occured per minute (**Fig.3.5 a**,

top row), the duration of each event (middle row), and the inter-event interval between neighboring events (bottom row). Significant increases in the number of events were noted on *Snap25 GCaMP6s*, whereby P21 the bulk of the number of events was around 5 per minute (one-way ANOVA age-F:2204.51 p=0). As expected, the median inter-event duration decreased with respect to age (one-way ANOVA age-F:1579.03 p=0). Unexpectedly, for the *Snap25 GCaMP6s* animals, I found that the average event duration did not change during the first two weeks of development, but finally decreased by P21 (one-way ANOVA age-F:471.01 p=0).

When I compared all mouse lines, I observed a progressive increase in the number of events per domain as the animals age (two-way ANOVA age and GCaMP driver, age-F:6918.08 p=0, driver-F:417.67 $p<0.0001$, interaction-F:569.65 p=0). Corresponding to this, the inter-event interval decreased with age for all mouse lines (two-way ANOVA age and GCaMP driver, age-F:882.34 p=0, driver-F:255.76 $p<0.0001$, interaction-F:80.91 $p<0.0001$). When analyzing the duration of each event in the layer specific lines, I observed a progressive decrease in the event duration for both the *Cux2-creER; Ai95* (middle row, right panel, blue) and the *Rbp4-Cre; Ai95* (orange) mice (two-way ANOVA age and GCaMP driver, age-F:336.31 $p<0.0001$, driver-F:4685.20 p=0, interaction-F:184.90 $p<0.0001$).

Utilizing the binarized time course, I calculated the regional percent of activation time and identified the cortical domains that were the most active in each mouse line. *Snap25 GCaMP6s* mice were most active at every age when compared to the *Cux2-CreER; Ai95* and *Rbp4-Cre; Ai95* mice (**Fig.3.5 b**). At each age the most active cortical areas shifted, where in *Snap25 GCaMP6s* animals the highest amount of activity was focused in the auditory cortex at P6, retrosplenial and higher order motor cortices at P10, and widespread higher order areas

at P14. *Cux2-CreER; Ai95* mice showed the highest activity in barrel and body core primary regions at P6, the most active cortical areas at P10 matched the *Snap25 GCaMP6s* pattern. They had equal activity across the whole cortex by P14. rbp4 tended to have low sparse activity, but higher percentage activations in primary cortical areas.

I characterized the percent of domains that were active with respect to the percentage time there was cortical activity, showing increasing activity based on mouse line and age (**Fig.3.5 c**). Each histogram shows the mean and standard deviation of the amount of time all animals spend in the binned cortical percentages, the average of each animal was shown by the red line (one-way ANOVA average values, age-F:24.53 $p < 0.0001$). Early *Snap25 GCaMP6s* mice (P2-P4) showed similar distributions, for the majority of time little to no domain activity was observed across the entirety of the cortex. By P5-P6, the cortex became more active, and the average progressively increased to closer to 50% of the time through P7-P9. By P10-P12, the activities were evenly distributed, with the majority and minority of activities across the cortex. During P13-P21, the percent of active areas shifted higher, with a long tail to no activity, maintaining that 50% of the cortex on average was active. TAM dosing (P6 green histogram) of Snap25-GCaMPs mice had no effect on the distribution of cortical activity when compared to the age matched non-dosed animal.

Cux2-CreER; Ai95 and *Rbp4-Cre; Ai95* mice showed decreased cortical activity when compared to age matched *Snap25 GCaMP6s*, but also showed increasing cortical activities with age alone (two-way ANOVA based on age and GCaMP driver, age-F:49.83 $p < 0.0001$, driver-F:77.02 $p < 0.0001$, interaction-F:4.08 $p < 0.05$). *Rbp4-Cre; Ai95* mice showed no significant shifts in activities with r age based on the post-hoc analysis. *Cux2-CreER; Ai95*

mice, however, showed an increase in the amount of activity between P6 and P10, however did not show significant increase from P10 to P14. TAM dosed *Snap25 GCaMP6s* did not show any difference from age-matched non-dosed mice.

Inter-hemispheric correlations

Development of functional long range cortical connectivity is pivotal for the two hemispheres to work in conjunction with each other. Visual inspection of the data indicated an apparently increased inter-hemispheric symmetry through developmental ages. To quantify this effect in an unbiased manner, I explored Pearson's correlation coefficients with hierarchical clustering and mapping similar inter-hemispheric positioned domains.

Hierarchical clustering showed a u-shaped developmental profile with high percentages of inter-hemispheric clustering in neonates animals (P2-P3) and adolescent animals (P14-P21). Given that hierarchical clustering depends on the distance between points in n-dimensional space (where n is the number of domains per animal), I explored the distribution of distances in each snap25 GCaMP6s mouse (**Fig.3.6 a**; top panel; black: maximal distance, grey:mean distance, light grey:40th percentile). Selection of cluster defining distance was shown as the dashed line (top), just below 40th percentile. This resulted in an increasing number of clusters (middle panel) which peaked at P9, followed by a decrease of clusters to P21 (one-way ANOVA age-F:12.05 $p > 0.0001$). When I queried which percentage of clustered domains had domains on both hemispheres, and found that P9 had the lowest percentage (bottom panel, one-way ANOVA age-F:10.78 $p < 0.0001$).

Given the high symmetry at P21, I inspected Pearson's coefficient of the two posi-

tional nearest domains on the contralateral hemisphere. After mapping the domains, I plotted the centroids of one hemisphere, colored based on the correlation coefficient across ages and mouse lines (**Fig.3.6 b**). In all *Snap25 GCaMP6s* lines I observed a high level of correlation in the body core region. The frontal motor cortex first developed high correlation values (P10-P14), followed by the sensory regions (P14-P21). In the *Cux2-CreER; Ai95* mice, I observed little correlation at the youngest age (P6), which ultimately became similar to the *Snap25 GCamP6s* line (P14). Of note, similar to the *Snap25-GCamp6s* mice, it was interesting that the binocular region of the primary visual cortex in the Cux2-CreER; Ai95 mice was highly correlated at P10. The *Rbp4-Cre; Ai95* mice showed similar symmetry as the *Cux2-CreER; Ai95* mice in the motor regions early on (P6, with seemingly increasing correlation through age), however it was more difficult to assess due to the reduced numbers of domains in older animals.

When analyzing the distribution of correlation values based on interhemispheric mapping, I observed a v-shaped developmental profile in *Snap25 GCamP6s* mice (one-way ANOVA age-F:116.80, $p < 0.0001$). There was a decrease in correlation values from P2 to P11, followed by a rapid increase in correlation through P21 (**Fig.3.6 c**; top panel) Both P6 *Cux2-CreER; Ai95* and *Rbp4-Cre; Ai95* animals showed significant decrease in correlation, ultimately achieving similar symmetries between P10 and P14 (bottom panel; cux2 : blue, rbp4: orange) (two-way ANOVA age and GCaMP driver, age-F:316.00 $p < 0.0001$, driver-F:74.07 $p < 0.0001$, interaction-F:29.63 $p < 0.0001$) There was no change in symmetry profile when *Snap25 GCaMP6s* non-dosed compared to TAM gavaged (green) mice.

Discussion

In this study I explored the early postnatal development of functional cortical activity patterns at mesoscale. I presented data from three different calcium based imaging mouse lines, one pan-neuronal including both excitatory and inhibitory neurons (*Snap25 GCaMP6s*) and two layer specific excitatory subpopulations of neurons (*Cux2-creER*/upper layers and *Rbp4-cre*/lower layers independently mated with *Ai95 floxed GCaMP6f* mouse line). Utilizing an ICA-pipeline I developed, I segmented the cortex into functional units at various ages of development [63]. This data-driven method reduced the number of time series that were needed for analysis, while maintaining the informative resolution of mesoscale. Wherein, I was able to assess sub-regional networks and how they interacted with each other across the majority of cortex.

Many different models have been proposed based on cortical columns and I find it difficult not to speculate that these developmental domains may be the closest units across the neocortex to date. Cortical columns were first proposed by Mountcastle, but subsequently integrated into several models of columnar organization. Many have discussed the cortical column with respect to the barrel fields and ocular dominance, as they are the most clearly separated components of their respective regions. However, cortical columns have been difficult to identify in other cortical regions due to the flow of information from one patch of cortex to the next. Several evidence indicate that the developmental domains I identified may be a close representation across the whole cortex. First, in our previous characterization of P21 recordings, I was able to show the subsequent recordings produced similar domain maps [63]. Second, between P6 to P9 in the snap25 GCaMP6s mice, the numbers of units did not change even though the

cortex was expanding, shown by the decrease in domain density (Figure 1). This is more surprising when considering that the complexity of the neural signal across the cortex increased through this time period shown by event detection activity (**Fig.3.5**) and hierarchical cluster analysis (**Fig.3.6**), yet ICA still identified the same number of components. Third, domain sizes between all three P6 mouse lines were very similar, and the domains had similar distribution of eccentricity and areas across the cortex (**Fig.3.2**). While more experimentation is still needed to validate this line of thinking, I am excited by the prospect.

These domains assist our understanding of the cortex based on regional constraints or to infer progressive modifications of structural circuitry based on age-related functional changes. I show that domains do not have the same uniform areas, but rather regional biases for larger domains that develop in the anterior-medial portions of the cortex. Further, I showed that the domain derived from functional activity maintained known regional borders, exemplified by the highly eccentric domains surrounding known regions (**Fig.3.2**). As the animal aged, I observed that these domains amalgamated into larger functional units that showed similar activity based on each independent spatial component.

Layer specific changes can also be observed in the developing cortex. The rapid domain growth in the Rbp4-Cre; Ai95 mice between P6 and P10 was striking, suggesting a rapid cortico-cortical connection maturation. It was reported that deeper fluorescent changes were associated with more diffuse signal as it passed through the upper layers [61], however I believe in this case in a strong case of circuit maturation. The same rate of change was not observed between P10 and P14, although more dendritic maturation was reported at this later ages [8, 10, 26]. Further, the Cux2-creER; Ai95 mice showed a drastic increase in domain size

between P10 and P14. This further highlighted the progressive development of each cortical layer that has been explored in the past [6, 26].

Due to the fact I was recording from population dynamics with calcium imaging techniques, I was well aware of the slow dynamics that I would be recording from, yet were struck with some of the frequency and event results. First, all three mouse groups showed an increase in frequency from P6 to P10, then returned to a slower frequency by P14 (**Fig.3.4**). While I was aware of each regional shift in frequency, large scale dynamic shifts across the cortex were unexpected. Further, the layer specific frequency and event progression advanced to our expectations, where the events increased in number while decreasing their duration (**Fig.3.5**). However, snap25 GCaMP6s did not follow this same pattern.

In fact, for the majority of postnatal ages I did not observe a decrease in event duration, but rather the events were sustained. I believe this was due to the diverse population of cells that expressed the calcium indicator, including all layers of excitatory and inhibitory neurons. This was also consistent with the predominant slow frequencies I observed during the first two postnatal weeks of development. As the circuit matures during the first two weeks, neurons get activated in non-uniform fashion, all the cell types extending out the event and slowing down the recorded frequency. This slow frequency and long event did not carry on to P21 animals, which indicated that the circuit needs further maturation into the third week of development.

Mesoscale imaging is a useful tool to use in conjunction with other experiments to explore more complicated circuit maturations. This body of work has demonstrated novel ways of thinking about these large scale imaging techniques and highlighted aspects that need further exploration. While the majority of the field is focused on their respective areas of interest, I

hope our study puts primary regions in perspective of the whole cortex functional development. Each primary region is not only behaving with their respective systems, but coinciding with functional activities across cortical domains at early ages.

Figures

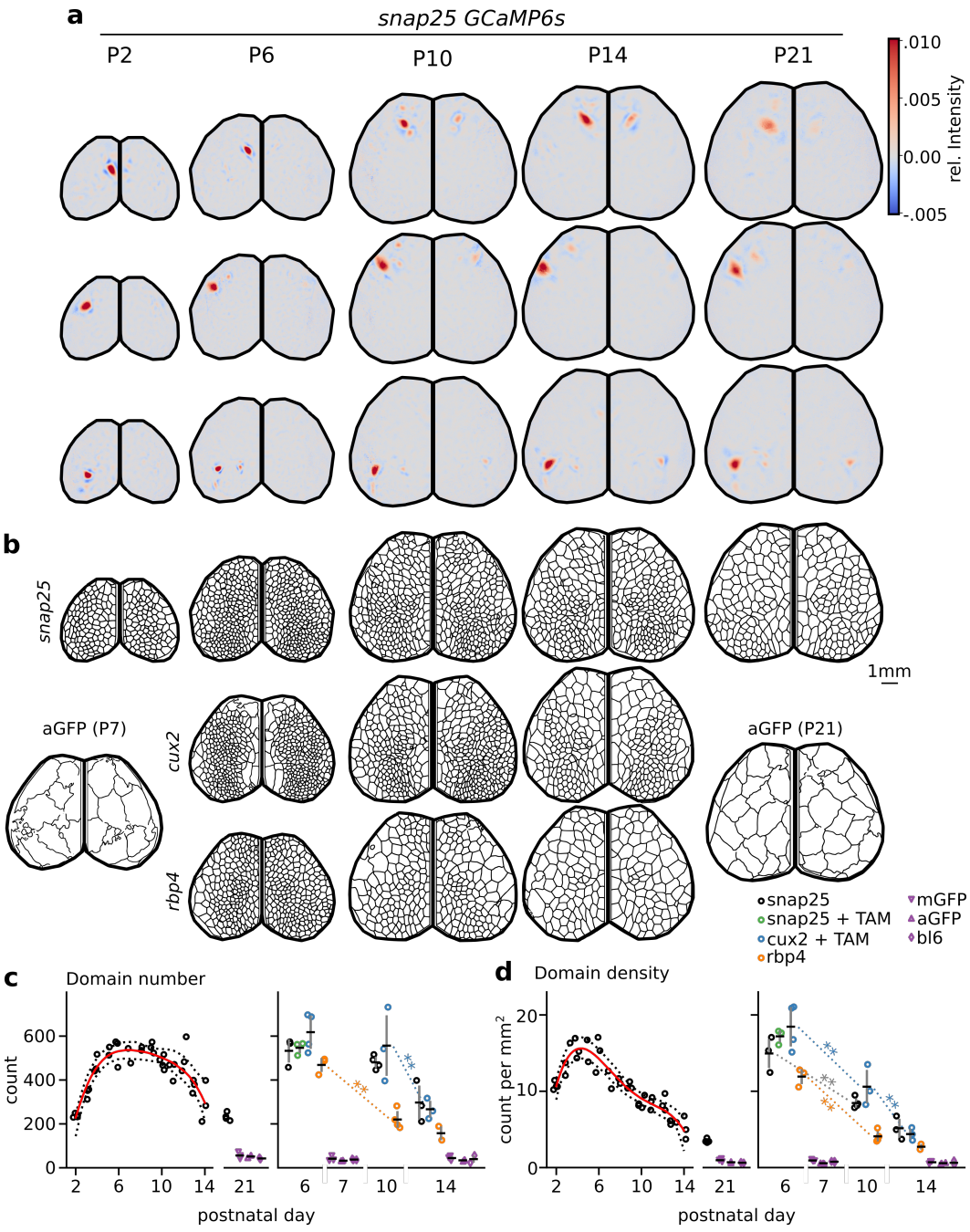

Figure 3.1: Domain maps created from eigenspace show an evolving development of cortical units based on cell types.

(a) Examples of eigenvectors of neural signals at postnatal day(P) 2, P6, P10, P14, and P21. Similar eigenvectors can be found at various stages of development and across multiple regions. Purported motor (top), head-sensory (middle), and visual (bottom) cortices. (b) Examples of domain maps created by max projection through the eigenspace for the same animals in Snap25-GCaMP6s mice (top row). Examples of domain maps created by data from layer specific driven GCaMP expression from upper layers (cux2; middle row) and layer 5 (rbp4; bottom row) data. P7 and P21 aGFP controls are also shown for comparison. (c) Domain number for all domain maps created for each age and condition across both hemispheres. (d) Domain density for all experiments takes the number of domains for each experiment and is divided by the sum of areas of all domains that comprise the cortex. snap25 GCaMP6s (black circles), snap25 GCaMP6s with TAM dose (green circles), cux2 creER; floxed GCaMP6f + TAM (blue); rbp4 cre; floxed GCaMP6f (orange); mGFP (purple down-triangle); aGFP (purple up-triangle), c57/bl6 mouse (purple diamond)

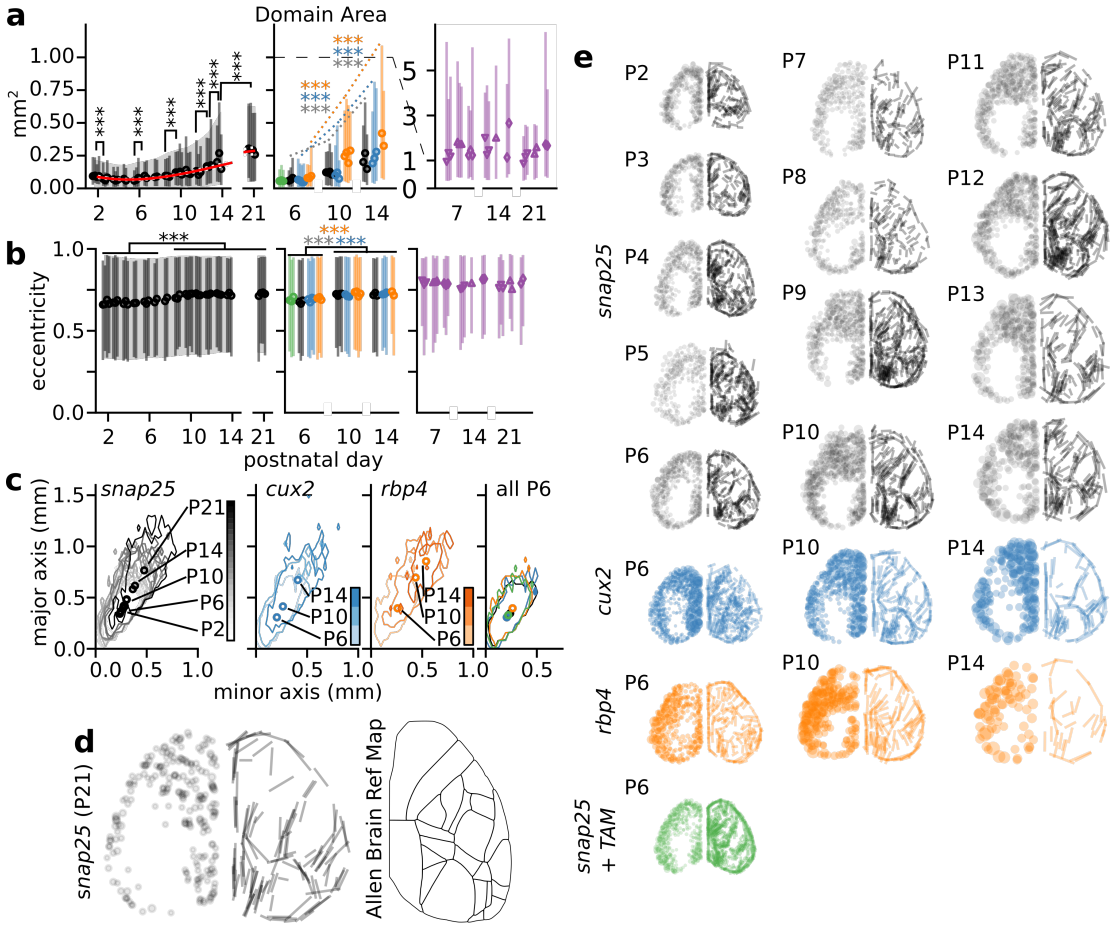

Figure 3.2: Domain sizes increase with age and shapes are not random. (a) Domain surface area plotted by age. Each individual animal is shown by their mean (open circle) and the 95th confidence interval (bar). Model was fit showing the progressive increase in area (red line, model error shown in white dashed lines). 95th confidence interval of the samples based on bootstrap shown (light grey filled). All genotypes and controls shown. (b) Eccentricity of domains plotted by animal, with 95th confidence and mean indicated (similar to a).(c) Contour of the full distribution of major and minor axes plotted by age. Gradient colors represent the age, light colors are young, decreasing in brightness to older ages. Far right panel overlaps the densities of all P6 data. (d) Upper quartiles of both surface area and eccentricity shown for P21 animal (left) with the adult Allen brain atlas map (right). All the centroids of each the highest size quartile shown (dots), plotted lateral to medial. Highest eccentricity quartile shown by plotting the major axis of each domain in its characterized orientation (lines). (e) Similar plotting as in (d) for each age and genotype. *Snap25 GCaMP6s* (black), *Snap25 GCaMP6s* with TAM dose (green), *Cux2-creER; Ai95* (blue); *Rbp4-cre; Ai95* (orange); mGFP (purple down-triangle); aGFP (purple up-triangle), c57/cl6 mouse (purple diamond).

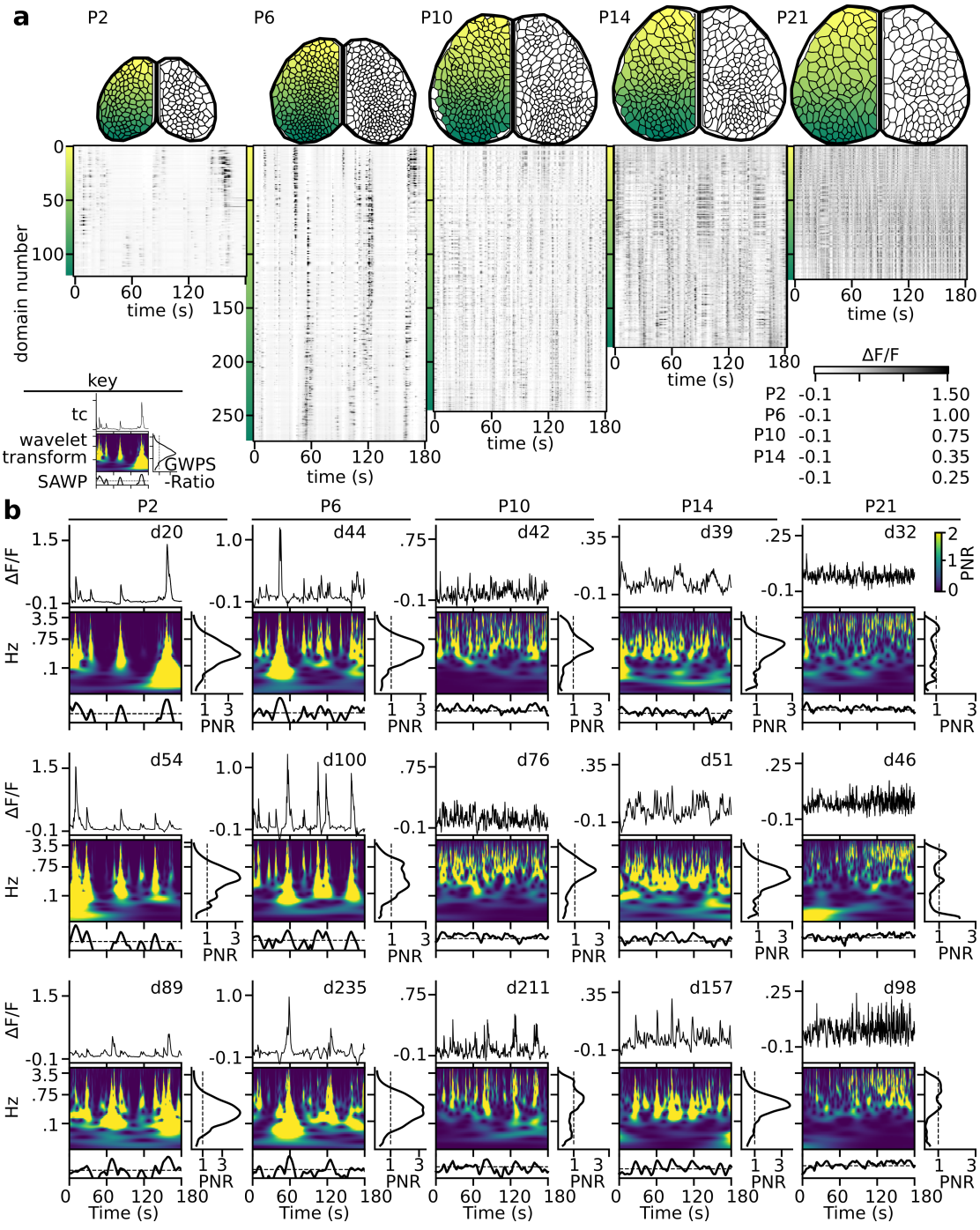

Figure 3.3: Extracted time courses from one hemisphere show progressive change in time course dynamics across ages.

(a) 3 minute examples of half domain datasets from P2, P6, P10, P14, and P21 snap25 GCaMP6s mice. Time courses plotted in change in fluorescence with respect to mean fluorescence (Δ F/F). All time courses from one hemisphere sorted by centroid position of each domain from anterior (yellow) to posterior (green). Under each domain map, all time courses from each domain are plotted (grey scale; lower limit kept at -0.1, variable upper limits based on age, see legend lower right). (b) Three examples per age of time course and the resulting wavelet transform ratio with red-noise normalization. Below the wavelet transform is the scaled average wavelet power (SAWP) through 0.1-5 Hz, showing when prominent wavelet activations occur (significance indicated values above the dashed line). The global wavelet power spectrum (GWPS) projected through the full time course exists to the right of the wavelet transform, indicating prominent frequencies above 95

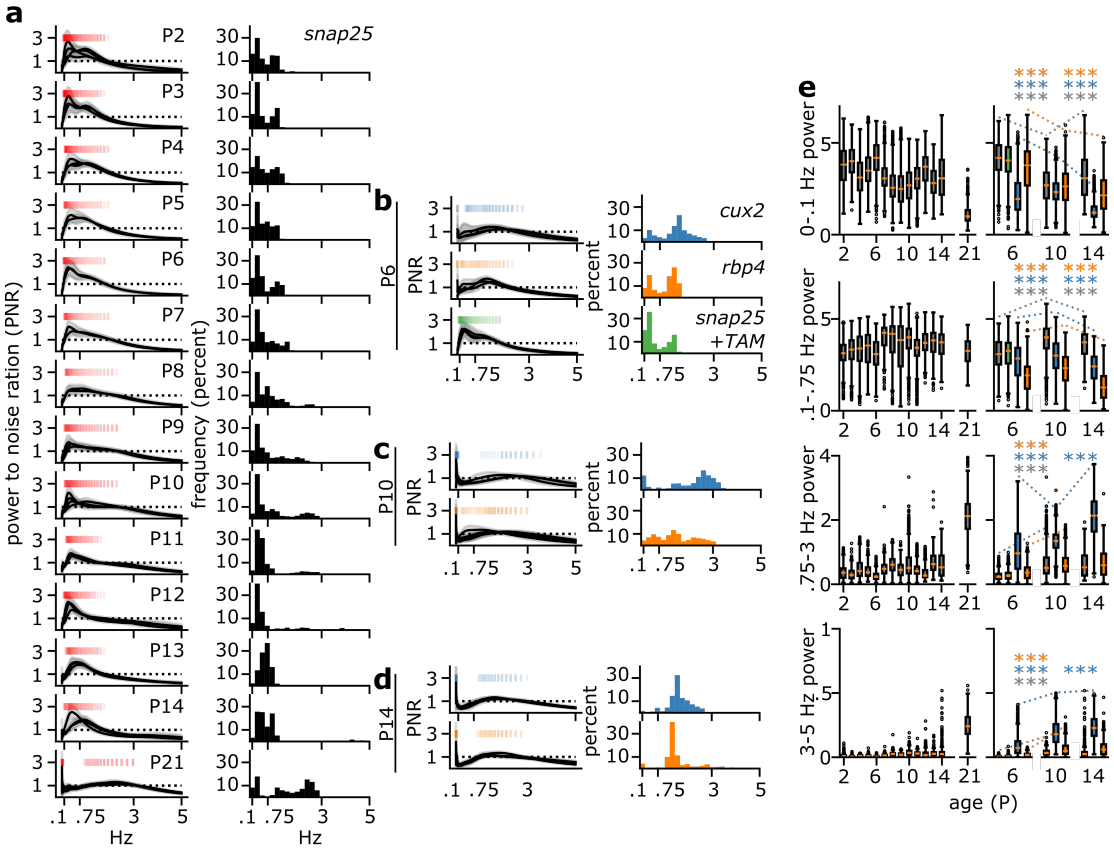

Figure 3.4: Prominent frequencies that arise at each developmental age. (a) Average GWPS of each animal (left column black line) with variance within each animal (left column grey), separated by age. Frequencies that are prominent in that age bracket, based on the mean GWPS, indicated (red) above the power ratio curves. Histograms of the local maxima of each domain GWPS (right column, black histogram). Similar analysis as (a) for P6 (b), P10 (c), and P14 (d) layer specific (cux2: blue; rbp4: orange) and snap25 GCaMP6s TAM control animals (green). (e) Integrated window power based on prominent frequencies seen in the histograms of local maxima. Top row corresponds to integration of all power between 0 and 0.1 Hz, second row 0.1-0.75 Hz, third row 0.75-3 Hz, and bottom row 3-5Hz. Left column shows power integration based on age, and the right column shows comparisons between the 4 conditions across each age assessed: *snap25 GCaMP6s* (grey), *snap25 GCaMP6s* + Tamoxifen (green), *cux2* (blue), and *rb4* (orange).

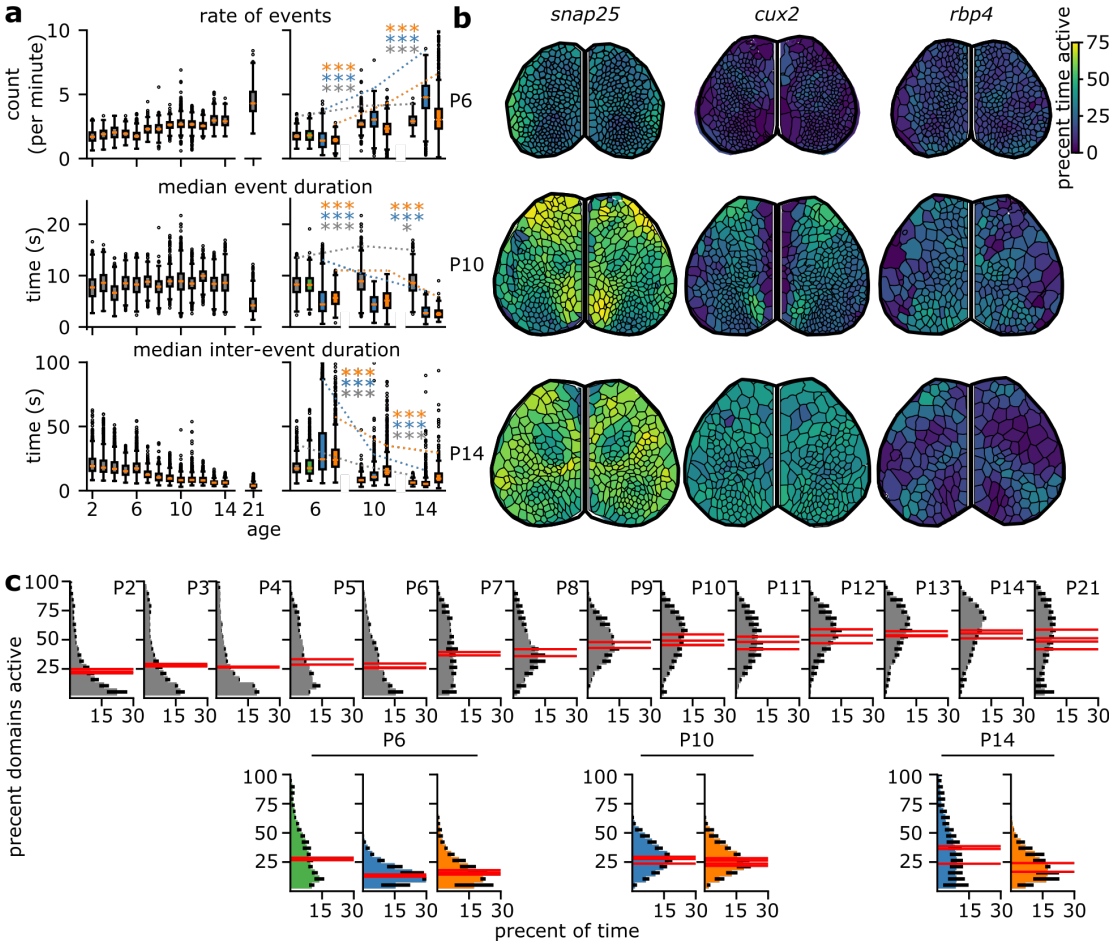

Figure 3.5: Wavelet analysis shows regional and age dependent activity patterns. (a) Box and whisker plots for the number of events, the duration and the inter-event duration for each domain. Rate of events per minute (top row), the event duration plotted per second (middle row), and the inter-event duration (bottom row) for *snap25 GCaMP6s* animals (grey), snap25 GCaMP6s + TAM (green), cux2 (blue), and rbp4 (orange) plotted by each age. (b) Example heatmaps for activity percentage for P6, P10, and P14 recordings of snap25, cux2, and rbp4. (c) Density plots showing the percentage of domains that are active (y-axis) at what percentage of time (x-axis) for each age and experimental condition. To show intra-animal variation, individual red lines are the average value for each animal in the dataset.

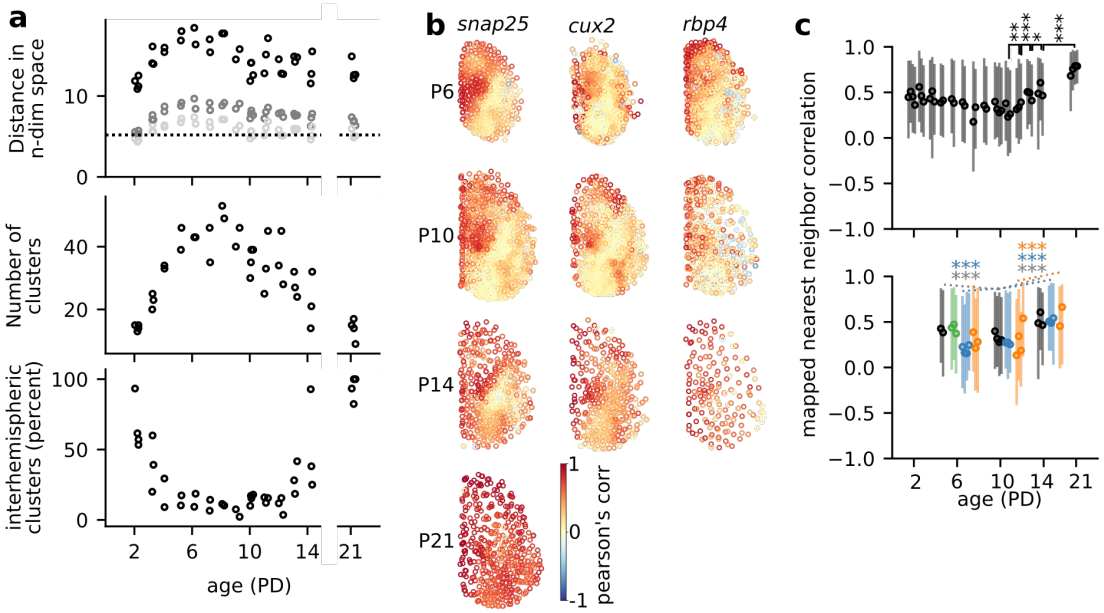

Figure 3.6: Interhemispheric symmetry drastically increases after P10. (a) Hierarchical cluster analysis based on Pearson's correlation coefficient. Top: distribution of the distances in n-dimensional space (n = number of domains). Maximal distance (black), mean distance (dark grey), 40th quantile (light grey). Dotted line indicates the defining distance for a cluster to be formed. Middle: The number of final clusters based on age and the defining distance. Bottom: The percentage of clusters that had components on both hemispheres. (b) Mapping each domain onto the opposite hemisphere, k-means clustering was used to determine the closest existing domain. Scatter plot showing one hemisphere, with color based on Person's correlation values for each pair of time courses (Blue: negative correlation, yellow: no correlation, red: positive correlation. (c) Range plots of each independent experiment (95th CI: bar; mean: Open circle) for *Snap25 GCaMP6s* (black), *Snap25 GCaMP6s* + TAM (green), *Cux2-cre; Ai95* (blue), and *Rbp4-cre; Ai95* (orange)

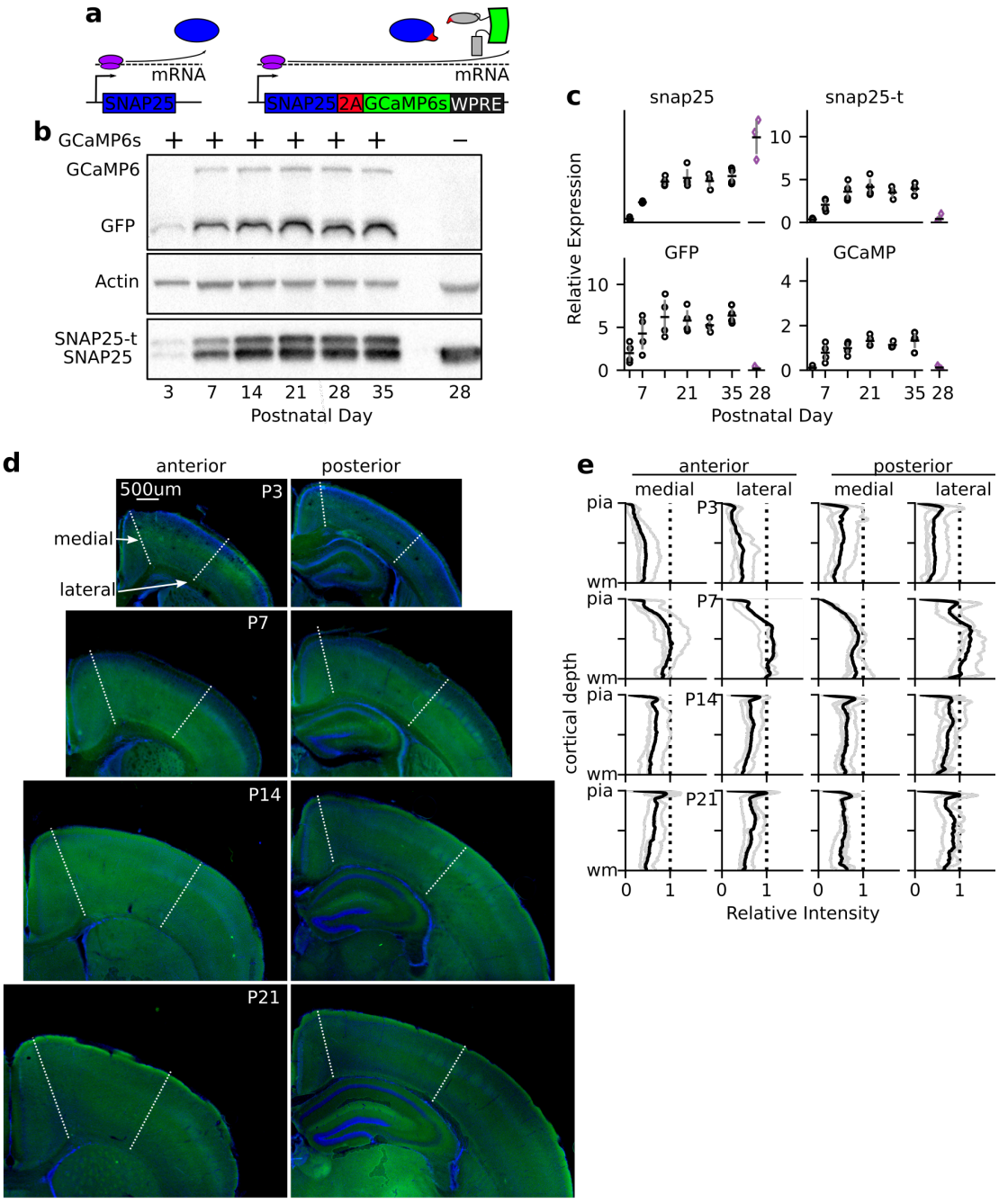

Figure 3.7: Snap25 driven expression of GCaMP increases during the first two weeks of development, then stabilizes.

(a) Genetic construct from derived mouse line (Madison et al. 2015). Self proteolytic cleavage between Snap25 and GCaMP6s proteins. (b) Western blot of Snap25 GCaMP animals increasing in age from left to right (P3, P7, P14, P21, P28, P35) and control data on the far right column (age P28). Genotypes indicated on top of each column, verified via PCR, indicating positive (+) for the transgene or not (−). Sizes (kDa) on the right of the gel based on the protein ladder. (c) Quantification of four bands from the western blot data, non-modified copy of Snap25 (upper left) and its transgenic analog (upper right), GFP (bottom left) and GCaMP6s (bottom right). (d) Histological expression of GCaMP6s in P3, P7, P14, and P21 animals, one anterior and posterior section were selected for analysis. GCaMP expression was measured in the medial portion of each section and a lateral section. (e) Quantification of the GCaMP6s at the same cortical locations in d. Individual animals are in grey and the mean of all animals is in black. Data was normalized to the highest average value.

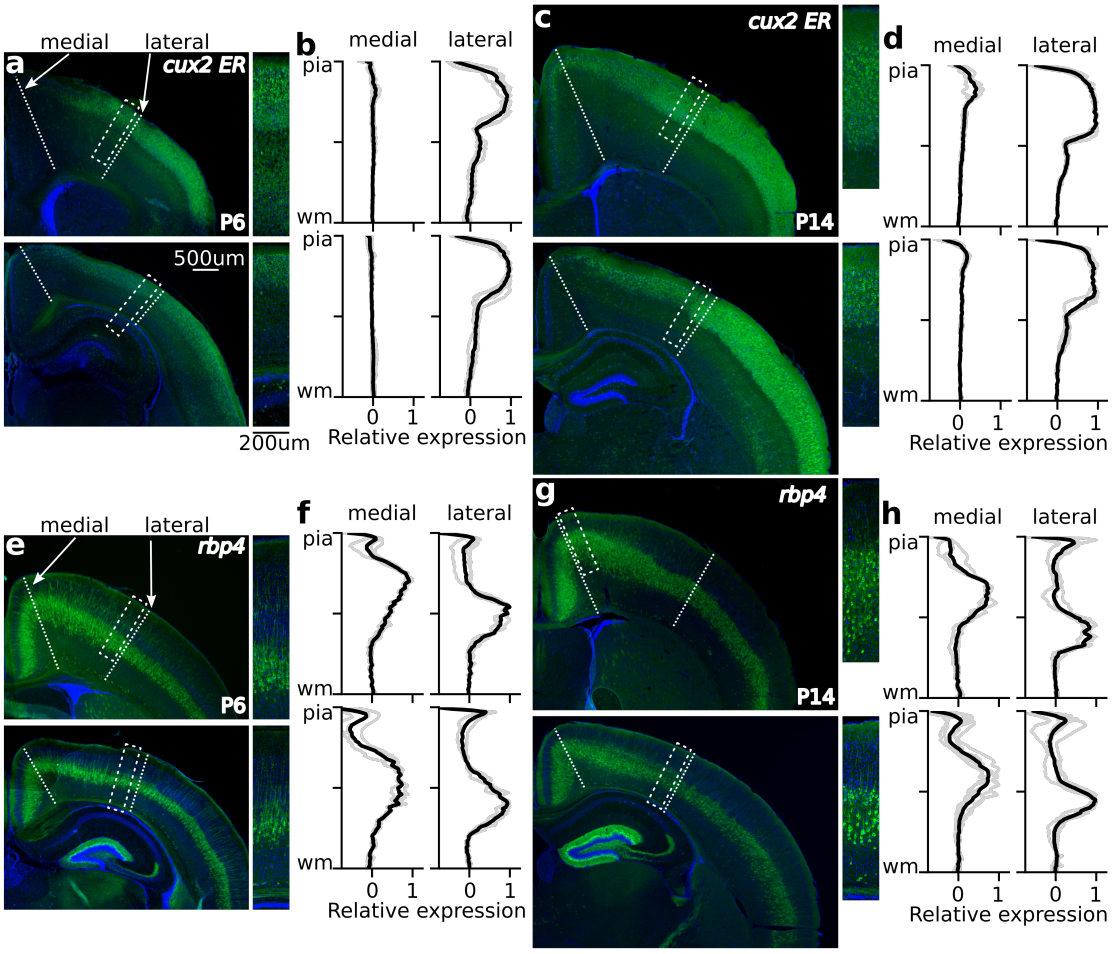

Figure 3.8: Immunostaining shows robust expression of GCaMP6s in layers 2-4 and in deep cortical layers in mice containing *Cux2-CreER* and *Rbp4-Cre* as early as P6. (a,c) Two example sections from *Cux2-creER; Ai95* taken at 2.5x and 10x magnification. White boxed region in low magnification is the location of the 10x image. The medial and lateral dashed lines indicate the areas of quantification. (e, f) Two example sections from *Rbp4-cre; Ai95* with similar sections as in (a). P6 (a,e) and P14 (c,g) are shown with their respective expression histograms (b,d,f,h), plotted from pia surface to the white matter (wm). Relative expression was calculated based on background (0) and the max intensity (1) of each section. Individual animals are shown in grey. Mean histogram is shown in black.

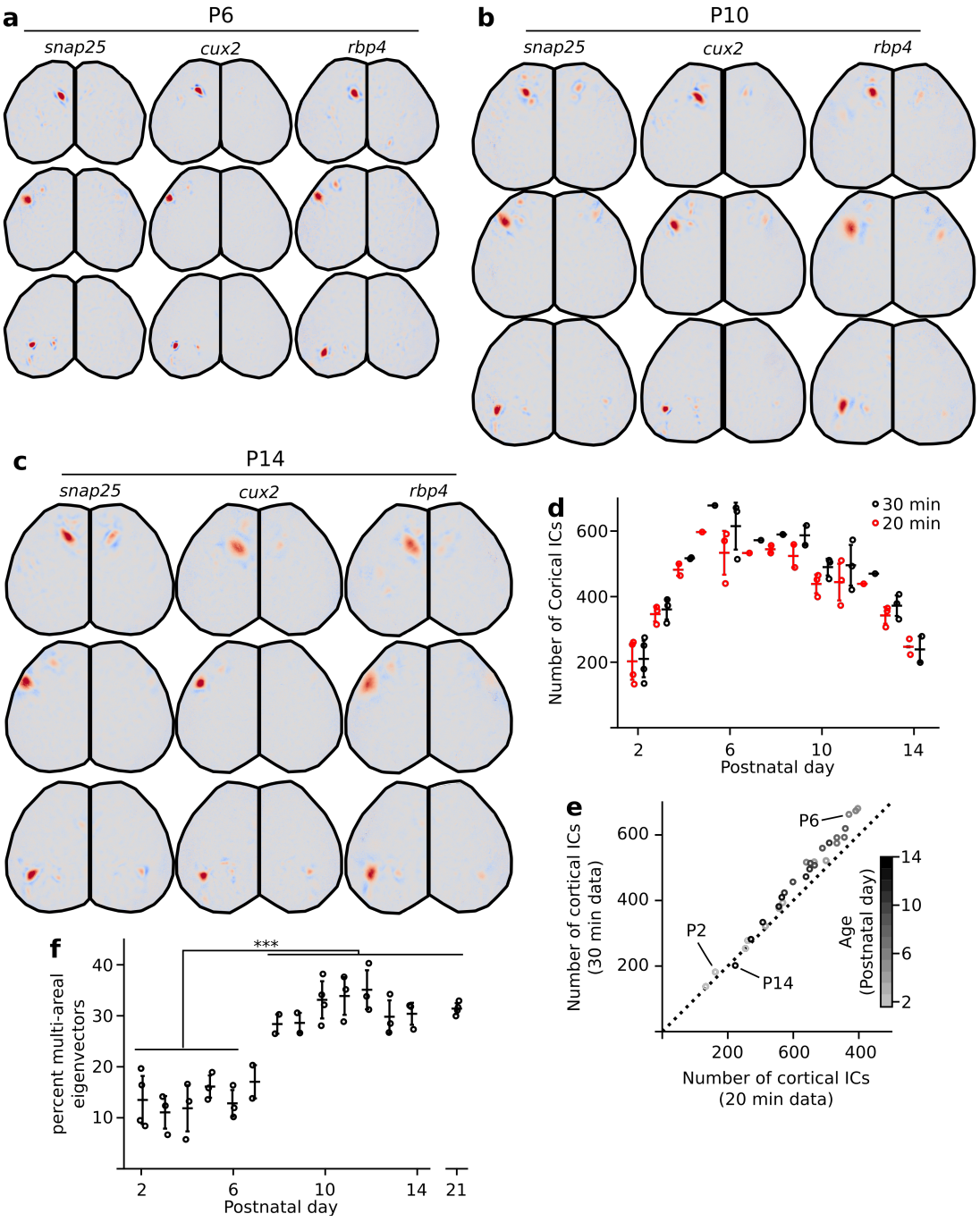

Figure 3.9: Example Eigen-decompositions from the three mouse lines, verification of maximal decomposition and increasing number of multi-areal eigenvectors.

Example eigenvectors from cux2, rbp4, and snap25 driven GCaMP mouse lines at ages (a) P6, (b) P10, and (c) P14. Snap25 are the same eigenvectors as Figure 1. (d,e) Verification maximal eigenvector decomposition, comparison of 20 min to 30 minutes worth of data. Increases in spatially complex animals, independent of temporal complexity. (f) Quantification of the percentage of multi-areal independent components. Significant increase in the number of multi-areal eigenvectors seen after the first week of development.

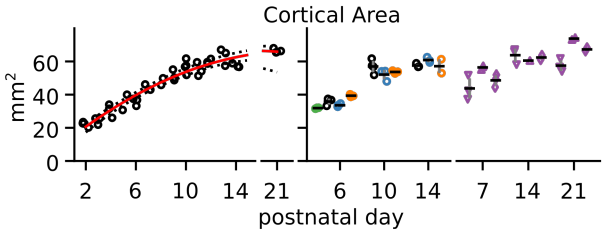

Figure 3.10: Cortical area defined by domain map increases with respect to age. (a) Total area of the cortex defined by the domain map for all GCaMP expressing mice (circles): *Snap25 GCaMP6s* (black), *Snap25 GCaMP6s* + TAM (green), *cux2-creER; Ai95* (blue), *Rbp4-cre; Ai95* (orange). Control (purple) for mGFP (down triangle), aGFP (up triangle), or C57/bl6 mice (diamond). High variance in the controls due to cortical areas that were not included in the domain map generation.

Methods

Mice

All animal studies were conducted in accordance with the UCSC Office of Animal Research Oversight and Institutional Animal Care and Use Committee protocols. All mice were maintained on C57/Bl6 background in UCSCs mouse facilities. P2-21 *Snap25 GCaMP6s* transgenic mice (JAX: 025111), P6-14 floxed GCaMP6f transgenic mice (Ai95; JAX:028865) crossed with either Rbp4-cre (MMRRC 031125-UCD) or Cux2-creER (graciously gifted to us by Dr Bin Chen [19]) strains were used for cortex specific activity. Cux2 creER were bred at least 10 generations onto the C57/Bl6 background. Control mice were P7-21 *Cx3cr1 GFP* (JAX: 005582), P7-21 *Aldh1 GFP* (MGI: 3843271), and P7-21 C57/Bl6 non-transgenic mice.

To identify *Snap25 GCaMPs* expressing mice, a single common forward primer (5'-CCC AGT TGA GAT TGG AAA GTG-3') was used in conjunction with either transgene specific reverse primer (5'-ACT TCG CAC AGG ATC CAA GA-3'; 230 band size) or control reverse primer (5'-CTG GTT TTG TTG GAA TCA GC-3'; 498 band size). The expression of this transgene resulted in pan-neuronal expression of GCaMP6s throughout the nervous system in adulthood [35]. To identify floxed GCaMP6f homozygous mice a mutant forward (5'-ACG AGT CGG ATC TCC CTT TG-3'; 450 band size) or internal control (5'-AAG GGA GCT GCA GTG GAG TA-3', 297 band size) were run with a common reverse primer (5'-CCG AAA ATC TGT GGG AAG TC-3'). The presence of CRE was determined via PCR using primers (5'-CAT CGC TCG ACC AGT TTA GT-3') and (5'-CGA TGC AAC GAG TGA TGA GG-3'; 375 band

size) To identify GFP expressing mice a forward (5'-CCT ACG GCG TGC AGT GCT TCA GC-3') and reverse (5'-CGG CGA GCT GCA CGC TGC GTC CTC-3'; 400 band size) PCR amplification was used to identify which animals had the GFP transgene.

Tamoxifen dosing

P1-2 neonates were orally gavaged at 75mg/kg/day Tamoxifen (Sigma) for three consequetive days from a 20mg/mL tamoxifen concentration in corn oil. Pups were allowed to rest a full day prior to imaging.

Surgical procedure

All mice were anesthetized with isoflurane (2.5% in pure oxygen) for the procedure. Body temperature was maintained at 37C for the duration of the surgery and recovery using a feedback-regulated heading pad. Lidocaine (1%) was applied subcutaneous on the scalp, followed by careful removal of skin above the skull. If the eyes were open, opthalmic ointment was used protect the eyes during the surgery. The cranium was attached using to two head bars using cyanoacrylate, one across the occipital bone of the skull and the other on the lateral parietal bone. Animals were allowed to recover for 1-1.5 hours to regain cognitive function. P2-8 were imaged on the heating pad, those that were expecially young (P2-P3) were further wrapped in a pulled cotton nest. After the mouse were ambulatory, mice were transferred to a rotating disk for the duration of the recording. At the end of each recording session, the animal was either euthanized or perfused and the brain dissected.

Recording calcium dynamics

In-vivo wide-field fluorescence recordings were collected in a minimally invasive manner. Imaging through the skull by single-photon excitation light from two blue LED light (470 nm; Thorlabs M470L3) produces a green fluorescent signal that is collected through coupled 50*mm* Nikon lenses (f5.6 / f1.2, optical magnification $\sim$ 1x) into a scientific CMOS camera (PCO Edge 5.5MP, 6.5μm pixel resolution). The top lens in the tandem lens setup was used to focus on the cortical surface, thereby lowering the magnification slightly; anatomical representation for each pixel corresponded to $6.9 \pm 0.2\mu$m (min: 6.7μm, max: 7.2μm). Excitation light was filtered with a 480/30 nm bandpass (Chroma Technology AT480/30x) and the emission signal was filtered with 520/36 nm bandpass (Edmund Optics 67-044). Data collection was performed in a dark, quiet room with minimal changes in ambient light or sound, thus the brain activity recorded was resting state without direct stimulation. Raw data was written directly as a set of 16 bit multi-image TIFF files.

The total amount of data recorded for each animal was generally at least 40 min and the amount of time in between video segments was less than 1 minute. All analyzed data of ages less than P14 consisted of three sequential full spatial resolution recordings concatenated together giving 30 min of data sampled at 10 frames per second for each video decomposition. P21 animals consisted of 20 min of data, with exact same sampling parameters as the younger mice[63].

ICA and Data processing

ICA decompositions were run on a single computing cluster node having 1024 GB of RAM. After ICA processing, selection of artifacts and domain map generation (all with a blur parameter of 51) was computed on local computers with 16-32GB of RAM. [63, 42]

Domain feature generation and clustering analysis

Scipy hierarchical clustering was utilized to determine cluster (scipy.cluster.hierarchy) Shape metrics for each domain were calculated based on regionprops (Skimage.measure.regionprops)

Wavelet Mean filtration

Wavelet decomposition on the time series signals were performed with a $\omega = 4$ morlet wavelet family, code adapted from C. Torrence and G. Compo,[55] available at

URL: http://paos.colorado.edu/research/wavelets/

Significance was determined using the 95th percentile of a red-noise model fit to the time series autocorrelation. Frequency distributions are all displayed as the ratio of the global wavelet specturm, relative to the noise cutoff. For wavelet filtering, the original signal was rebuilt excluding all frequency signals in a certain range.

Immunofluorescence

Free floating tissue sections of 30 m coronal neocortex were blocked with 5% normal donkey serum (NDS), then incubated with chicken anti-GFP (Novus; 1:2000 dilution) at 4C for 24-48 hours.

Secondary antibody, Goat anti-chicken (1:500) conjugated to Alexa488, was applied for 4 hours at room temperature and tissue was counterstained with DAPI. Images collected on Zeiss AxioImager as tiled images at 2.5x magnification and z-stack at 10x utilizing the Apotome feature.

Western blot

P3, P7, P14, P21, P28, and P35 full cortices of Snap25 GCaMP6s mice with control non-transgenic liter mates were dissected and flash-frozen to be stored at 80C. Cortices were homogenized using sonification (Diagenode) in lysis buffer containing 50 mM Tris/HCl, 0.25% (w/v) sodium deoxycholate, 150 mM NaCl, 1 mM EDTA, 1% (w/v) Nonidet P40 (Igepal), with phosphatase inhibitor cocktail (Thermo Halt) containing AEBSF, aprotinin, bestatin, E-64, leupeptin and pepstatin A.

Then 10 g samples were resolved on 10% (w/v) mini-SDS polyacrylamide gels (Bio-Rad) and transferred to PVDF membranes. Ponseau stain was used to verify protein load and help approximate cutting of the PVDF membranes into two sections, along with the protein ladder (Amersham Pharmacia Biotech) (upper section with GFP, GCaMP and actin, lower with snap25). Membranes were blocked with 5% (w/v) non-fat dried skimmed milk in PBS-T [17 mM KH2PO4, 50 mM Na2HPO4, 1.5 mM NaCl, pH 7.4, and 0.05% (v/v) Tween 20] for 1 h at room temperature and then incubated 48 hrs at 4C in PBS-T plus 5% (w/v) non-fat dried skimmed milk containing primary rabbit anti--actin (1:5000; Novus) added to the upper membrane and rabbit anti-snap25 (1:5000 dilution; Novus) antibodies added to the lower. Protein bands were detected using an ECL kit (Amersham Pharmacia Biotech), with horseradish peroxidase-conjugated secondary Donkey anti-rabbit antibodies (1:30000,Novus). Upper mem-

brane was then striped of antibodies using acidic (pH 2.2) aqueous solution containing 0.1% SDS and 1% Tween20. Re-processing of the upper membrane was then done utilizing rabbit anti-GFP (1:5000 dilution; Sigma Aldrich) and the same secondary antibody concentrations.

Relative protein levels in each sample were determined by comparison with -actin levels. Relative intensities of the protein bands were quantified by scanning densitometry using ImageJ.

Regression analysis (with model error)

Polynomial fit for each data point was conducted. The degree of the polynomial was selected by minimizing the error of the fit, while assessing each coefficient for its impact on the model. An ANOVA was run on the coefficients to determine how each coefficient differed from zero and each other.

Once the degree was selected, bootstrapping was utilized to determine model error (1000 iterations of 1000 randomly selected points for domain metric fitting; 1000 iterations of 25 randomly selected points for animal metrics)

Sample confidence interval calculation

95% confidence interval was determined via bootstrapping 1000 iterations of 300 samples spanning the day before and after of development. Confidence intervals were reported for each iteration, resulting distributions of lower and upper confidence intervals. The mean values of the distributions were reported.

Statistical significance

Statistical significance was calculated using OLS models from `statsmodel.formula.api` with Holm-Sidak multiple testing correction ($p \leq 0.5$: *; $p \leq 0.01$: **; $p \leq 0.001$: ***). Model significance is determined by the F-statistic, and significance of two-group analyses (p>|t|) are calculated with t-tests.

Chapter 4

Conclusion

In this dissertation, I developed tools necessary to explore functional activity across the full dorsal cortical sheet of the first three weeks of developing neonatal mice. This big-data approach showcases the the strengths of collecting sufficient spatial representation with goals of understating functional activity. Further, this work begins to take on some of the challenging tasks of developmental neuroscience. Such that the end product shows that functional cortical activity patterns give key provisional insight into postnatal brain development. These insight can be capitalized on in future experimentation involved with cortical maturation and circuit formation.

Past work in the lab

Previous work in the lab that culminated to this body of work, show data-driven methods to spatially segment the cortex that represents underlying cortical structure. Independent component analysis (ICA) decomposes the functional calcium recordings into components, comprised of a

spatial eigenvector and functional activity time course. We can maximally project through the spatial representation to achieve a domain map to discretize the cortex, while maintaining the underlying structural limitations based on function. This reduces the number of time courses from a few million of individual pixels down to a few hundred domains. These domain maps can be replicated on from subsequent recordings with slight variations [63].

Machine learning in identification of Neural components

In the Chapter 2, I demonstrate what constitutes neural components from the ICA decomposition. With the use of controls, I was able to identify many of the common non-GCaMP based signals that are produced while recording from non-anesthetized freely behaving mice. These included movement and vascular artifact components that I saw in many of recordings. However, I also found many of the hemodynamic components in these non-GCaMP controls that were very distinct from tight globular components found from GCaMP animals. These neural components were unmixed from any vascular influence, a common trait in all our data collections that was never seen in control animals.

From each of the components, I was able to extract key features and teach a machine to identify each neural component with 97% accuracy. Spatial features were key to identifying each neural component. Selecting on the neural components allows for a filtered representation of only the GCaMP6s signals without the influence of motion or vascular artifacts. Upon establishing the machine learning algorithm, I was able to show high accuracy, even on novel data that was not used to generate the model.

Developmental profiles of functioning developing cortex

In chapter 3, I utilized the ICA-pipeline across development in three different mouse lines to establish functional developmental profiles. The pan-neuronal (*snap25 GCaMP6s*) was utilized to establish a full functional profile, wherein I showed developmental day-resolution changes for the first two weeks. The upper (*Cux2-creER; Ai95*) and lower cortical layer (*Rbp4-cre; Ai95*) were used to compare and understand how varying subpopulations contributed to the signal seen in the pan neuronal line. We were impressed with the fact that each of these lines produced similar looking components, strengthening our conviction that ICA components are based on structurally relevance.

In each mouse line, I found that the domain maps across development give insight in underlying circuit development. We consistently found that highly eccentric domains outline major structural divisions between regions. Further, I found in all mouse lines that the anteromedial portion of the cortex contained larger domains than the sensory regions of the cortex. The highest density of cortical domains occurs between P4-P6 in our *Snap25 GCaMP6s* line. The same number is maintained through P9 even while the cortex is growing, thereby decreasing in density. We found that around P10-P14, the circuits assemble into larger functional marco-circuits, contributing to the larger domains seen in older animals.

When I analyzed the time courses from each of these domains, I showed that pan-neuronal expression of calcium indicators maintain slower frequencies than the subpopulation dynamics, but eventually speed up by P21. Both layer specific subpopulations (lower and upper layers) have faster dynamics, with increased counts of shortened events. We show that the

structured activity in pan neuronal developing cortex begins to shift at P5, increasing in activity until P9. Many of the higher order regions that specialize in sensory-sensory and sensory-motor integration begin to show high levels of activity at the end of the first week, which increases as the animals ages.

Inter-hemispheric correlation occurs rapidly after P11, first in the frontal cortex then in sensory regions posterolateral cortex. Day to day significant increases are shown after the P11 date. Lower layer subpopulations did not have much change in correlation during the first week, but showed significant increase from P10 to P14. Upper layer neurons showed ever increasing cortical inter-hemispheric connections.

Applications and future possibilities

Our ambitions are that our work will benefit the greater community and enable higher quality and robust data collection and analysis. Given the excitement and positive feedback for our data acquisition and analytical pipeline that I received from several researchers, I have developed a number of collaboration. These collaborations will produced a number of papers to further our understanding of how cortical patterns are affected by behavior, genetics or environment.

Two such collaboration with the Zuo lab at UCSC look at how environmental conditions change cortical circuits. The first collaboration investigated how cortical patterns change in adults after chronic stress and are ameliorated after treatment with a novel therapeutic [32]. This work benefited from our filtration methods identifying only neural signal imaged through the cranial window, independent of vascular artifacts. Through this analysis, we were able to

compare calcium imaging between the mesoscale cortical activity with microscale 2-photon single cell imaging, to elucidate the therapeutic benefits of the novel drug. The second collaboration with the Zuo lab investigated cortical changes due to perinatal penicillin exposure across the entire dorsal cortex [47]. Utilizing the domain maps from our pipeline, we were able to show widespread lengthening of functional events due to perinatal exposure to penicillin, with little effect on regional coherence.

In addition to collaborating with neuroscientists, I have worked to strengthen our statistical understanding and collaboration. I have completed a designated emphasis in statistics and had various rigorous discussions of analytic approaches with Raquel Prado from the UCSC statistics department. These discussions have helped shape the way I approach the analysis and conclusions I attain from this large dataset. Further, the Prado lab has made advances in Bayesian approaches to model complex time series data [71], of which they actively working toward an efficient calculation and better time series models for the data I have produced. Development of these models will only further our understanding of developing cortical dynamics.

Our hope and desire is that this work inspires developmental neuroscientists in pursuing previously untapped and difficult investigations. Future work can use this technology in order to map regional networks independent of a reference map. As such, researcher will be able to create their own maps regardless of the genetic manipulations that resulted in aberrant maps. This opens up the possible investigation into cortico-cortical connections that have eluded researcher for years. Of course, we will always further our work with the promise to share our open-source code and hardware development for the benefit of all.

Bibliography

[1] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, 2014.

[2] James B. Ackman, Timothy J. Burbridge, and Michael C. Crair. Retinal waves coordinate patterned activity throughout the developing visual system. *Nature*, 490:219 – 225, 2012.

[3] James B. Ackman, Hongkui Zeng, and Michael C. Crair. Structured dynamics of neural activity across developing neocortex. *bioRxiv*, Dec 2014.

[4] A Agmon, LT Yang, DK O'Dowd, and EG Jones. Organized growth of thalamocortical axons from the deep tier of terminations into layer iv of developing mouse barrel cortex. *Journal of Neuroscience*, 13(12):5365–5382, 1993.

[5] Camille Allène, Adriano Cattani, James B. Ackman, Paolo Bonifazi, Laurent Aniksztejn, Yehezkel Ben-Ari, and Rosa Cossart. Sequential generation of two distinct synapse-driven network patterns in developing neocortex. *Journal of Neuroscience*, 28(48):12851–12863, 2008.

[6] Paul G. Anastasiades and Simon J. B. Butt. A role for silent synapses in the development of the pathway from layer 2/3 to 5 pyramidal cells in the neocortex. *Journal of Neuroscience*, 32(38):13085–13099, 2012.

[7] Charles T Anderson, Patrick L Sheets, Taro Kiritani, and Gordon M G Shepherd. Sublayer-specific microcircuits of corticospinal and corticostriatal neurons in motor cortex. *Nature Neuroscience*, pages 739–744, 2010.

[8] Kevin J. Bender, Juliana Rangel, and Daniel E. Feldman. Development of columnar topography in the excitatory layer 4 to layer 2/3 projection in rat barrel cortex. *Journal of Neuroscience*, 23(25):8759–8770, 2003.

[9] Ursula Boschert, Celestine O'Shaughnessy, Robin Dickinson, Michela Tessari, Caterina Bendotti, Stefan Catsicas, and Emilio Merlo Pich. Developmental and plasticity-related differential expression of two snap-25 isoforms in the rat brain. *Journal of Comparative Neurology*, 367(2):177–193, 1996.

[10] Ingrid Bureau, Gordon M.G Shepherd, and Karel Svoboda. Precise development of functional and anatomical columns in the neocortex. *Neuron*, 42(5):789–801, 2004.

[11] Daniel L. Chao, Le Ma, and Kang Shen. Transient cellcell interactions in neural circuit formation. *Nature Reviews Neuroscience*, 4:262–271, 2009.

[12] Adam S. Charles, Benjamin Falk, Nicholas Turner, Talmo D. Pereira, Daniel Tward, Benjamin D. Pedigo, Jaewon Chung, Randal Burns, Satrajit S. Ghosh, Justus M. Kebschull, William Silversmith, and Joshua T. Vogelstein. Toward community-driven big open brain

science: Open big data and tools for structure, function, and genetics. *Annual Review of Neuroscience*, 43(1):441–464, 2020. PMID: 32283996.

[13] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, Vivek Jayaraman, Loren L. Looger, Karel Svoboda, and Douglas S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, July 2013.

[14] Yachi Chen and Anirvan Ghosh. Regulation of dendritic development by neuronal activity. *Journal of Neurobiology*, 64(1):4–10, 2005.

[15] Mattia Chini and Ileana L. Hanganu-Opatz. Prefrontal cortex development in health and disease: Lessons from rodents and humans. *Trends in Neurosciences*, 44(3):227–240, 2021.

[16] Shen-Ju Chou, Zoila Babot, Axel Leingärtner, Michele Studer, Yasushi Nakagawa, and Dennis D. M. O'Leary. Geniculocortical input drives genetic distinctions between primary and higher-order visual areas. *Science*, 340(6137):1239–1242, 2013.

[17] Matthew T. Colonnese, Anna Kaminska, Marat Minlebaev, Mathieu Milh, Bernard Bloem, Sandra Lescure, Guy Moriette, Catherine Chiron, Yehezkel Ben-Ari, and Rustem Khazipov. A Conserved Switch in Sensory Processing Prepares Developing Neocortex for Vision. *Neuron*, 67(3):480–498, August 2010.

[18] James C Dooley and Mark S Blumberg. Developmental 'awakening' of primary motor cortex to the sensory consequences of movement. *eLife*, 7:e41841, dec 2018.

[19] Santos J Franco, Cristina Gil-Sanz, Isabel Martinez-Garay, Ana Espinosa, Sarah R Harkins-Perry, Cynthia Ramos, and Ulrich Müller. Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science*, 337:746–749.

[20] Olga Garaschuk, Jennifer Linn, Jens Eilers, and Arthur Konnerth. Large-scale oscillatory calcium waves in the immature cortex. *Nature Neuroscience*, 3:452–459, 2000.

[21] A. Géron. Hands-on machine learning with scikit-learn and tensorflow. 2017.

[22] Joshua I. Glaser, Ari S. Benjamin, Roozbeh Farhoodi, and Konrad P. Kording. The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, 175:126 – 137, 2019.

[23] Luciano Custo Greig, Mollie B. Woodworth, Maria J. Galazo, Hari Padmanabhan, and Jeffrey D. Macklis. Molecular logic of neocortical projection neuron specification, development and diversity. *Nature Reviews Neuroscience*, pages 755–769, 2013.

[24] Ileana L. Hanganu, Yehezkel Ben-Ari, and Rustem Khazipov. Retinal waves trigger spindle bursts in the neonatal rat visual cortex. *Journal of Neuroscience*, 26(25):6728–6736, 2006.

[25] Julie A Harris, Karla E Hirokawa, Staci A Sorensen, Hong Gu, Maya Mills, Lydia L Ng, Phillip Bohn, Marty Mortrud, Benjamin Ouellette, Jolene Kidney, Kimberly A Smith, Chinh Dang, Susan Sunkin, Amy Bernard, Seung Wook Oh, Linda Madisen, and Hongkui Zeng. Anatomical characterization of cre driver mice for neural circuit mapping and manipulation. *Front Neural Circuits*, 8:76, 2014.

[26] Ryan J. Kast and Pat Levitt. Precision in the development of neocortical architecture: From progenitors to cortical networks. *Progress in Neurobiology*, 175:77–95, 2019.

[27] R. Khazipov, M. Minlebaev, and G. Valeeva. Early gamma oscillations. *Neuroscience*, 250:240–252, 2013.

[28] Rustem Khazipov, Anton Sirota, Xavier Leinekugel, Gregory L Holmes, Yehezkel Ben-Ari, and Gyrgy Buzski. Early motor activity drives spindle bursts in the developing somatosensory cortex. *Nature*, 432:758–761, 2004.

[29] Sergei Kirischuk, Anne Sinning, Oriane Blanquie, Jenq-Wei Yang, Heiko J Luhmann, and Werner Kilb. Modulation of neocortical development by early neuronal activity: Physiology and pathophysiology. *Front Cell Neurosci*, 11:379, 2017.

[30] Lynette Lim, Da Mi, Alfredo Llorca, and Oscar Marn. Development and functional diversification of cortical interneurons. *Neuron*, 100(2):294–313, 2018.

[31] Guillermina López-Bendito and Zoltán Molnár. Thalamocortical development: how are we going to get there? *Nature Reviews Neuroscience*, 4:276 – 289.

[32] Ju Lu, Michelle Tjia, Brian Mullen, Bing Cao, Kacper Lukasiewicz, Sajita Shah-Morales, Sydney Weiser, Lindsay P. Cameron, David E. Olson, Lu Chen, and Yi Zuo. An analog of psychedelics restores functional neural circuits disrupted by unpredictable stress. *Molecular Psychiatry*, pages 1476–5578, 2021.

[33] Heiko J. Luhmann. Review of imaging network activities in developing rodent cerebral cortex in vivo. *Neurophotonics*, 4(3):1 – 8, 2016.

[34] Ying Ma, Mohammed A. Shaik, Sharon H. Kim, Mariel G. Kozberg, David N. Thibodeaux, Hanzhi T. Zhao, Hang Yu, and Elizabeth M. C. Hillman. Wide-field optical mapping of neural activity and brain haemodynamics: considerations and novel approaches. *Philosophical Transactions of the Royal Society B*, 371:20150360, 2016.

[35] Linda Madisen, AleenaR. Garner, Daisuke Shimaoka, AmyS. Chuong, NathanC. Klapoetke, Lu Li, Alexander vanderBourg, Yusuke Niino, Ladan Egolf, Claudio Monetti, Hong Gu, Maya Mills, Adrian Cheng, Bosiljka Tasic, ThucNghi Nguyen, SusanM. Sunkin, Andrea Benucci, Andras Nagy, Atsushi Miyawaki, Fritjof Helmchen, RuthM. Empson, Thomas Knöpfel, EdwardS. Boyden, R.Clay Reid, Matteo Carandini, and Hongkui Zeng. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron*, 85(5):942–958, 2015.

[36] Andre Marques-Smith, Daniel Lyngholm, Anna-Kristin Kaufmann, Jacqueline A. Stacey, Anna Hoerder-Suabedissen, Esther B.E. Becker, Michael C. Wilson, Zoltán Molnár, and Simon J.B. Butt. A transient translaminar gabaergic interneuron circuit connects thalamocortical recipient layers in neonatal somatosensory cortex. *Neuron*, 89(3):536–549, 2016.

[37] Marat Minlebaev, Matthew Colonnese, Timur Tsintsadze, Anton Sirota, and Roustem Khazipov. Early gamma oscillations synchronize developing thalamus and cortex. *Science*, 334(6053):226–229, 2011.

[38] Hidenobu Mizuno, Tomoo Hirano, and Yoshiaki Tagawa. Evidence for activity-dependent

cortical wiring: Formation of interhemispheric connections in neonatal mouse visual cortex requires projection neuron activity. *Journal of Neuroscience*, 27(25):6760–6770, 2007.

[39] Hidenobu Mizuno, Koji Ikezoe, Shingo Nakazawa, Takuya Sato, Kazuo Kitamura, and Takuji Iwasato. Patchwork-type spontaneous activity in neonatal barrel cortex layer 4 transmitted via thalamocortical projections. *Cell Reports*, 22:123 – 135, 2018.

[40] Zoltán Molnár, Shuji Higashi, Richard Adams, and Keisuke Toyama. Earliest interactions between thalamus and cortex. pages 47–79, 2001.

[41] Zoltán Molnár, Guillermina López-Bendito, Juan Small, L. Donald Partridge, Colin Blakemore, and Michael C. Wilson. Normal development of embryonic thalamocortical connectivity in the absence of evoked synaptic activity. *Journal of Neuroscience*, 22(23):10313–10323, 2002.

[42] Brian R Mullen, Sydney C Weiser, Ascencio Desiderio, and James B Ackman. Automated classification of signal sources in mesoscale calcium imaging. *bioRxiv*, 2021.

[43] Shigehiro Namiki, Hiroaki Norimoto, Chiaki Kobayashi, Kei Nakatani, Norio Matsuki, and Yuji Ikegaya. Layer iii neurons control synchronized waves in the immature cerebral cortex. *Journal of Neuroscience*, 33(3):987–1001, 2013.

[44] L Paninski and JP Cunningham. Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Current Opinion in Neurobiology*, 50:232 – 241, 2018. Neurotechnologies.

[45] Linden Parkes, Ben Fulcher, Murat Yücel, and Alex Fornito. An evaluation of the efficacy,

reliability, and sensitivity of motion correction strategies for resting-state functional mri. *NeuroImage*, 171:415 – 436, 2018.

[46] Anna A Penn and Carla J Shatz. Brain waves and brain wiring: The role of endogenous and sensory-driven neural activity in development. *Pediatric Research*, 45:447 – 458, 1999.

[47] James Perna, Brian Mullen, Taohui Liu, Michelle Tjia, Sydney Weiser, Ju Lu, James Ackman, and Yi Zuo. Perinatal penicillin exposure affects cortical development and adolescent sensory processing. *In Review*, 2021.

[48] Gabrielle Pouchelon, Frdric Gambino, Camilla Bellone, Ludovic Telley, Ilaria Vitali, Christian Lscher, Anthony Holtmaat, and Denis Jabaudon. Modality-specific thalamocortical inputs instruct the identity of postsynaptic l4 neurons. *Nature*, 511:471–474, 2014.

[49] Gerald R. Prescott and Luke H. Chamberlain. Regional and developmental brain expression patterns of snap25 splice variants. *BMC Neuroscience*, 12:35, 2011.

[50] Raimon H.R. Pruim, Maarten Mennes, Daan van Rooij, Alberto Llera, Jan K. Buitelaar, and Christian F. Beckmann. Ica-aroma: A robust ica-based strategy for removing motion artifacts from fmri data. *NeuroImage*, 112:267 – 277, 2015.

[51] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville. Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Processing Magazine*, 30(3):58–70, 2013.

[52] Simon Rumpel, Gunnar Kattenstroth, and Kurt Gottmann. Silent synapses in the immature visual cortex: Layer-specific developmental regulation. *Journal of Neurophysiology*, 91(2):1097–1101, 2004. PMID: 14762153.

[53] Swathi Srivatsa, Srinivas Parthasarathy, Zoltán Molnár, and Victor Tarabykin. Sip1 downstream effector ninein controls neocortical axonal growth, ipsilateral branching, and microtubule growth and stability. *Neuron*, 85(5):998–1012, 2015.

[54] Rodrigo Suárez, LauraR. Fenlon, Roger Marek, Lilach Avitan, Pankaj Sah, GeoffreyJ. Goodhill, and LindaJ. Richards. Balanced interhemispheric cortical activity is required for correct targeting of the corpus callosum. *Neuron*, 82(6):1289–1298, 2014.

[55] Christopher Torrence and Gilber. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meterological Society*, 79:61–78, 1998.

[56] SebnemN. Tuncdemir, Brie Wamsley, FloorJ. Stam, Fumitaka Osakada, Martyn Goulding, EdwardM. Callaway, Bernardo Rudy, and Gord Fishell. Early somatostatin interneuron connectivity mediates the maturation of deep layer cortical circuits. *Neuron*, 89(3):521–535, 2016.

[57] Matthieu P. Vanni, Allen W. Chan, Matilde Balbi, Gergely Silasi, and Timothy H. Murphy. Mesoscale mapping of mouse cortex reveals frequency-dependent cycling between distinct macroscale functional modules. *Journal of Neuroscience*, 37:7513–7533, 2017.

[58] Mai-Anh T. Vu, Tülay Adalı, Demba Ba, György Buzsáki, David Carlson, Katherine Heller, Conor Liston, Cynthia Rudin, Vikaas S. Sohal, Alik S. Widge, Helen S. Mayberg,

Guillermo Sapiro, and Kafui Dzirasa. A shared vision for machine learning in neuroscience. *Journal of Neuroscience*, 38(7):1601–1607, 2018.

[59] Chun-Lei Wang, Lei Zhang, Yang Zhou, Jing Zhou, Xiu-Juan Yang, Shu-min Duan, Zhi-Qi Xiong, and Yu-Qiang Ding. Activity-dependent development of callosal projections in the somatosensory cortex. *Journal of Neuroscience*, 27(42):11334–11342, 2007.

[60] Philip Washbourne, Peter M. Thompson, Mario Carta, Edmar T. Costa, James R. Mathews, Guillermina López-Bendito, Zoltán Molnár, Mark W. Becher, C. Fernando Valenzuela, L. Donald Partridge, and Michael C. Wilson. Genetic ablation of the t-snare snap-25 distinguishes mechanisms of neuroexocytosis. *Nature Neuroscience*, 5:19–26, 2002.

[61] Jack Waters. Sources of widefield fluorescence from the brain. *eLife*, 9:e59841, 2020.

[62] Xiaoxi Wei1, Neil Thomas, Nan E. Hatch, Min Hu, and Fei Liu. Postnatal craniofacial skeletal development of female c57bl/6ncrl mice. *Front. Physiol.*, 8, 2017.

[63] Sydney C. Weiser, Brian R. Mullen, Desiderio Ascencio, and James B. Ackman. Data-driven filtration and segmentation of mesoscale neural dynamics. *bioRxiv*, 2021.

[64] Aaron T. Winder, Christina Echagarruga, Qingguang Zhang, and Patrick J. Drew. Weak correlations between hemodynamic signals and ongoing neural activity during the resting state. *Nature Neuroscience volume*, 20:17611769, 2017.

[65] Fong Kuan Wong, Kinga Bercsenyi, Varun Sreenivasan, Adrián Portalés, Marian Fernández-Otero, and Oscar Marín. Pyramidal cell regulation of interneuron survival sculpts cortical networks. *Nature*, 557:668–673, 2018.

[66] Dongsheng Xiao, Matthieu P Vanni, Catalin C Mitelut, Allen W Chan, Jeffrey M LeDue, Yicheng Xie, Andrew CN Chen, Nicholas V Swindale, and Timothy H Murphy. Mapping cortical mesoscopic networks of single spiking cortical or sub-cortical neurons. *eLife*, 6:e19976, 2017.

[67] Benyi Xiong, Anan Li, Yang Lou1, Shangbin Chen, Ben Long, Jie Peng, Zhongqin Yang, Tonghui Xu, Xiaoquan Yang, Xiangning Li, Tao Jiang, Qingming Luo, and Hui Gong. Precise cerebral vascular atlas in stereotaxic coordinates of whole mouse brain. *Front. Neuroanat.*, 11, 2017.

[68] Nobuhiko Yamamoto and Guillermina López-Bendito. Shaping brain connections through spontaneous neural activity. *European Journal of Neuroscience*, 35(10):1595–1604, 2012.

[69] Jenq-Wei Yang, Shuming An, Jyh-Jang Sun, Vicente Reyes-Puerta, Jennifer Kindler, Thomas Berger, Werner Kilb, and Heiko J. Luhmann. Thalamic Network Oscillations Synchronize Ontogenetic Columns in the Newborn Rat Barrel Cortex. *Cerebral Cortex*, 23(6):1299–1316, 05 2012.

[70] Jenq-Wei Yang, Ileana L Hanganu-Opatz, Jyh-Jang Sun, and Heiko J Luhmann. Three patterns of oscillatory activity differentially synchronize developing neocortical networks in vivo. *J Neurosci*, 29:9011–9025.

[71] Wenjie Zhao and Raquel Prado. Efficient bayesian parcor approaches for dynamic modeling of multivariate time series. *Journal of Time Series Analysis*, 41(6):759–784, 2020.

[72] James Q. Zheng and Mu-ming Poo. Calcium signaling in neuronal motility. *Annual Review of Cell and Developmental Biology*, 23(1):375–404, 2007. PMID: 17944572.

[73] Jun Zhuang, Lydia Ng, Derric Williams, Matthew Valley, Yang Li, Marina Garrett, and Jack Waters. An extended retinotopic map of mouse cortex. *eLife*, 6:e18372, 2017.