

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Decoding Amplified targeted Transcripts with Fluorescent in situ Hybridization (DART-FISH) reveals cellular heterogeneity of human cerebral cortex in space

Permalink

<https://escholarship.org/uc/item/3wr8v61h>

Author

Chen, Chien-Ju

Publication Date

2023

Supplemental Material

<https://escholarship.org/uc/item/3wr8v61h#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Decoding Amplified targeted Transcripts with Fluorescent in situ Hybridization
(DART-FISH) reveals cellular heterogeneity of human cerebral cortex in space

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

Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Chien-Ju Chen

Committee in charge:

Professor Kun Zhang, Chair

Professor Eran Mukamel, Co-Chair

Professor Jerold Chun

Professor Kevin King

Professor Bing Ren

2023

Copyright

Chien-Ju Chen, 2023

All rights reserved.

The Dissertation of Chien-Ju Chen is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE	iii
TABLE OF CONTENTS	iv
LIST OF SUPPLEMENTARY FILES	v
LIST OF ABBREVIATIONS	vii
LIST OF FIGURES	x
LIST OF TABLES	xi
ACKNOWLEDGEMENTS	xii
ABSTRACT OF THE DISSERTATION	xvi
CHAPTER 1: INTRODUCTION	1
Anatomy and functions of human cerebral cortex	1
Development of human cerebral cortex	2
Arealization of cerebral cortex	5
Transcriptomic cytoarchitecture of human cerebral cortex	9
Technologies to measure gene expression in space	11
Outline of the first part of the dissertation	15
Outline of the second part of the dissertation	16
CHAPTER 2: DEVELOPMENT OF DECODING AMPLIFIED TARGETED TRANSCRIPTS WITH FLUORESCENT IN SITU HYBRIDIZATION(DART-FISH)	18
Introduction	18
DART-FISH workflow	19
Benchmarking and validation	25
Organization of cell types in the human primary motor cortex	30
Discussion	36
Acknowledgement for Chapter 2	37
CHAPTER 3: DART-FISH REVEALS CELLULAR HETERO- GENEITY OF HUMAN CEREBRAL CORTEX	39

Introduction	39
Transcriptomic profiling of single-cells in human cortical areas	41
Spatially resolved cell atlas of human cerebral cortex	50
Discussion	60
Acknowledgement for Chapter 3	63
CHAPTER 4: CONCLUSION AND PERSPECTIVES	64
MATERIALS AND METHODS	67
Reagents and enzymes	67
Post-mortem human brain tissue	67
Padlock probe production	68
DART-FISH sample preparation	70
DART-FISH data processing	73
RNAscope	78
Cell annotation	80
SUPPLEMENTAL FIGURES	83
REFERENCES	93

LIST OF SUPPLEMENTAL FILES

1. Chen_BICCN-DART-FISH-protocol.pdf
2. Chen_DART-FISH-paper-protocol.pdf
3. Chen_Enzymatic-padlock-probe-preparation.pdf
4. Chen_Supplementary Table 1.xlsx
5. Chen_Supplementary Table 2.xlsx

LIST OF ABBREVIATIONS

General terms

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

mRNA: Messenger RNA

rRNA: Ribosomal RNA

NGS: Next-generation sequencing

RNA-seq: RNA-sequencing

scRNA-seq: Single-cell RNA sequencing

snRNA-seq: Single-nucleus RNA sequencing

FISH: Fluorescence in situ hybridization

smFISH: single-molecule fluorescence in situ hybridization

osmFISH: Cyclic-ouroborous smFISH

DART-FISH: Decoding amplified targeted transcripts with fluorescence in situ hybridization

FOV: Field of view

RCA: Rolling circle amplification

rolony: RCA colony

SNR: Signal-to-noise ratio

NA: Numerical aperture

pcw: Post-conception week

BMP: Bone morphogenetic protein

Shh: Sonic hedgehog

FGF: Fibroblast growth factor

AIBS: Allen Institute for Brain Science

BICCN: Brain Initiative Cell Census Network

PFA: Paraformaldehyde

PA: Polyarcylamide

MRI: Magnetic resonance imaging

Brain terminology

VZ: Ventricular zone

SVZ: Subventricular zone

OSVZ: Outer subventricular zone

CP: Cortical plate

MZ: Marginal zone

vRG: Ventricular radial glia

oRG: Outer radial glia

tRG: truncated radial glia

NSC: Neural stem cell

IPC: Intermediate progenitor cell

MGE: Medial ganglionic eminence

CGE: Caudal ganglionic eminence

POA: Preoptic area

ACC: Anterior cingulate cortex

DFC: Dorsolateral prefrontal cortex

PFC: Prefrontal cortex

FI: Frontal insular cortex

M1C: Primary motor cortex

S1C: Primary somatosensory cortex

MTG: Middle temporal gyrus

A1C: Primary auditory cortex

AnG: Angular gyrus

V1C: Primary visual cortex

IT: Intratelencephalic

NP: Near-projecting

CT: Corticothalamic

Astro: Astrocyte

Endo: Endothelial cell

VLMC: Vascular and leptomeningeal cell

Oligo: Oligodendrocyte

OPC: Oligodendrocyte progenitor cell

Micro/PVM: Microglia/perivascular macrophage

WM: White matter

AP: Apical progenitor

LIST OF FIGURES

2.1	DART-FISH workflow: Decoding Amplified targeted Transcripts with Fluorescence in situ Hybridization	24
2.2	Benchmarking DART-FISH on the human motor cortex	29
2.3	DART-FISH mapping cell types in human M1 cortex	34
3.1	DART-FISH and SNARE-seq2 assays on 9 cortical areas	42
3.2	Nissl staining of 9 human cortical areas along rostrocaudal axis	45
3.3	Spatial distribution of excitatory neurons in 9 areas of human cerebral cortex	52
3.4	Spatial distribution of inhibitory neurons in 9 areas of human cerebral cortex	53
3.5	Spatial distribution of non-neuronal cells in 9 areas of human cerebral cortex	54
3.6	Proportion of cell types at class and subclass level across 9 cortical areas	56
3.7	E:I ratio across 9 cortical areas along the rostrocaudal axis	58
3.8	UMAPs of excitatory neurons, inhibitory neurons, and non-neuronal cells across 9 cortical areas	60
S1	Probe production strategy	83
S2	TBU gel check and size selection of padlock probes	85
S3	Gel embedding after reverse transcription, spot calling, and empty probabilities	86
S4	RiboSoma staining, spot assignment, and RNAscope validation	88
S5	Cell annotation of human M1C	89
S6	Summary and annotation of SNARE-seq nuclei across 9 areas of human cerebral cortex	90
S7	Cross-areal consensus taxonomy across 9 areas of human cerebral cortex	91

S8	Composition of 24 subclasses across 9 areas of human cerebral cortex	92
----	--	----

LIST OF TABLES

3.1	Deep sampling of areal nuclei by SNARE-seq2	44
3.2	Marker genes to annotate 24 cell subclasses of DART-FISH cells.	46
3.2	Marker genes to annotate 24 cell subclasses of DART-FISH cells (Continued).	47
3.2	Marker genes to annotate 24 cell subclasses of DART-FISH cells (Continued).	48
3.3	Annotation of DART-FISH cells in each cortical area	49
S1	Metadata of human brain donors	68
S2	TBE-Urea denaturing gel (TBU gel) casting solution recipe . .	70

ACKNOWLEDGEMENTS

I want to acknowledge many people including, my advisor Dr. Kun Zhang, my committee members, my parents, all members in the Zhang lab, collaborators, and friends for their help and support during my graduate study. First, I'd like to thank my PhD advisor, Dr. Kun Zhang, for taking me into his lab. I was in a bad shape when I switched to the Zhang lab at the end of my second year. He fosters a very open-minded lab environment and gives everyone his patience so that people have some room to fail and learn lessons from the failures. In Zhang lab, I had privileges to work with many talented students and scientists. I learned lots of principles of tool development with his mentorship.

Also, he loops me into his BICCN U01 project, giving me chances to apply our spatial transcriptomic tool, DART-FISH, to study human cerebral cortex and make some contributions to the scientific community. In this large-scale collaborative work, I ran across my committee members, Dr. Bing Ren and Dr. Jerold Chun, who offer a wealth of knowledge about epigenetics and neuroscience. I started gaining my knowledge about chromatin accessibility data analysis and knowledge about detailed neuroanatomy and brain development from our monthly U01 meetings. Besides, I want to thank my committee members, Dr. Eran Mukamel and Dr. Kevin King, who provided great suggestions in the mid of my graduate study and offered great career advice.

Besides, I want to thank my family and lab members for all their support and suggestions. First, I am grateful to my parents for always being so supportive to pursue high education. Without their support, I wouldn't even have the chance to study abroad and start this PhD study. Also, I'd like to thank all members in the Zhang lab, where I learned interesting science and intriguing ideas. Particularly, I want to thank Kian Kalhor and Mahsa Nifisi, who closely worked with me during the DART-FISH tool development. We shared our experimental tips and experiences of computational tools together and shouldered each other when we ran across failures

and depression. Ultimately, we pushed this work to the finish line and publish it on bioRxiv. The DART-FISH tool development is demonstrated in Chapter 2.

In addition, I am grateful for all collaborators in all research projects that I was involved during my PhD journey. First, I want to acknowledge our postdoc fellow, Dr. Jinghui Song, who worked closely with me in the U01 project. She not only offered phenomenal intellectual input but also gave me invaluable suggestions about practicing good scientific ethics and searching postdoc positions. Second, Dr. Carter Palmer in Dr. Jerold Chun's lab offered great support during mass data generation phase in the U01 project. He was so cooperative and patient to provide more samples whenever something went wrong. This U01 work is still ongoing when I draft this dissertation. Part of the work is demonstrated in Chapter 3. Third, I want to thank Dr. Van Ninh in Dr. Kevin King's lab. In our collaboration, I had the chance to apply DART-FISH to measure interferon stimulated genes in the infarct borderzones of mouse heart tissue, exemplify a good hypothesis-driven biological study. Besides, she also gave me good career advice and help me rehearse my seminar talk in preparation for the postdoc interview.

Together, I feel lucky and grateful to run across these outstanding people during my PhD study. I wouldn't have a chance to share my work in this dissertation without their advice, encouragement, and support.

Citations for the published materials

I was involved in the DART-FISH protocol optimization and performed experiments on human brain tissue sections. Also, I wrote the paper manuscript that constitute the majority part of Chapter 2 with Kian Kalhor and Kun Zhang. Besides, I generated all DART-FISH datasets, aligned and analyzed the datasets in Chapter 3. In addition, the contents in the MATERIALS AND METHODS section are mostly followed the paper manuscript used in Chapter 2 with slight modifications.

Chapter 2 is based on the following publication:

Kalhor, Kian*; Chen, Chien-Ju*; Lee, Ho Suk; Cai, Matthew; Nafisi, Mahsa; Que, Richard; Palmer, Carter; Yuan, Yixu; Zhang, Yida; Song, Jinghui; Knoten, Amanda; Lake, Blue B.; Gaut P. Joseph; Keene, Dirk; Lein, Ed; Kharchenko, Peter V.; Chun, Jerold; Jain, Sanjay; Fan, Jian-Bing; and Zhang, Kun. 2023. Mapping Human Tissues with Highly Multiplexed RNA in situ Hybridization, bioRxiv (under review by Nature Communications).

I am the co-first author of this paper and the primary author of Chapter 2.

Chapter 3 is based on the following work which will be published in the future:

Chen, Chien-Ju; Song, Jinghui; Palmer, Carter; Zemke, Nathan; Yang, Bing; Diep, Dinh; Lake, Blue B.; Zhang, Yida; Kharchenko, Peter V.; Ren, Bing; Chun, Jerold; and Zhang, Kun. BICCN work under preparation for publication.

I am the primary author of Chapter 3.

MATERIALS AND METHODS is mainly based on the following publication with slight modifications:

Kalhor, Kian*; Chen, Chien-Ju*; Lee, Ho Suk; Cai, Matthew; Nafisi, Mahsa; Que, Richard; Palmer, Carter; Yuan, Yixu; Zhang, Yida; Song, Jinghui; Knoten, Amanda; Lake, Blue B.; Gaut P. Joseph; Keene, Dirk; Lein, Ed; Kharchenko, Peter V.; Chun, Jerold; Jain, Sanjay; Fan, Jian-Bing; and Zhang, Kun. 2023. Mapping Human Tissues with Highly Multiplexed RNA in situ Hybridization, bioRxiv (under review by Nature Communications).

I am the co-first author of this paper and the primary author of the MATERIALS AND METHODS section.

ABSTRACT OF THE DISSERTATION

Decoding Amplified targeted Transcripts with Fluorescent in situ Hybridization
(DART-FISH) reveals cellular heterogeneity of human cerebral cortex in space

by

Chien-Ju Chen

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2023

Professor Kun Zhang, Chair

Professor Eran Mukamel, Co-Chair

Human cerebral cortex is tremendously complex with distinct functional areas. These functionally discrete areas lie on a large cortical sheet that has diverse cell types, cytoarchitectures, and transcriptomic variations. Previously, lots of efforts were spent to deciphering the areal cellular organizations and their molecular signatures. Recent studies that use single-cell/nucleus RNA-sequencing to profile the whole transcriptomes of human cortical areas extensively categorize the distinct transcriptomic cell

types in these areas but the spatial information was lost during cell/nucleus isolation. Thus, a spatial transcriptomic tool is required to further uncover the cytoarchitectures and their spatial transcriptomes.

Recently, many spatial transcriptomic tools have emerged but have different aspects of disadvantages such as low sensitivity, low spatial resolution, and low throughput. These limitations restrict their direct application to postmortem human brain tissue that have large size, general low tissue quality, and high autofluorescence signals. Here, we developed Decoded Amplified taRgeted Transcripts with Fluorescent in situ Hybridization or DART-FISH, a versatile padlock probe-based technology capable of profiling hundreds of genes in centimeter-sized human tissue sections at cellular resolution. As proof of principle, we measured 121 genes in human primary motor cortex (M1C), benchmarked, and validated its reproducibility, sensitivity, and specificity.

In addition, we applied DART-FISH to measure spatial expression of 484 genes in 9 human cortical areas along the rostrocaudal axis. Meanwhile, SNARE-seq2 was used to profile the whole transcriptomes of isolated nuclei from these areas. With the alignment of DART-FISH cells with pre-annotated SNARE-seq2 nuclei, spatially resolved transcriptomic cell atlases of human cerebral cortex were generated. In line with previous studies in M1C, these cortical areas are all composed of 24 cell subclasses of excitatory neurons, inhibitory neurons, and non-neuronal cells. Also, we found the cytoarchitectures of these cortical areas are similar except the rostral most anterior cingulate cortex (ACC) and the caudal most primary visual cortex (V1C). V1C showed the most drastic specialization, with tremendous expansion of layer 4 and increase of L4 IT neurons. To sum up, our findings suggest that these cortical areas have subtle cytoarchitectural and transcriptomic variations, except those areas at the poles of the rostrocaudal axis.

CHAPTER 1: INTRODUCTION

Anatomy and functions of human cerebral cortex

Neuroanatomy of human cerebral cortex

Human cerebral cortex is the outer layer of human cerebrum and contains diverse cell types organized in complicated structures. At the cellular level, it has been reported that the human brain contains 86 billion neurons and approximately 10 times more glial cells[1]. Strikingly, the cerebral cortex alone makes up to 82% of the mass and contain 16 billion neurons (19% of neurons in the adult human brain)[1]. In addition, cerebral cortex plays critical roles in sensory, motor, cognitive functions, and intellectual capabilities. Anatomically, human cerebral cortex can be further divided into neocortex (isocortex) and allocortex. Allocortex includes olfactory cortex, hippocampus, and paleocortex. It is phylogenetically older than neocortex and characterized by less layers of cell in these structures. On the other hand, neocortex is characterized by the 6 laminar layers of neurons and non-neuronal cells[2]. It undergoes tremendous expansion and folding during evolution and constitutes 90% of the cerebral cortex. It's widely considered it's the expanded and convoluted neocortex in cerebral cortex that set human beings apart from rodents (mostly *Mus musculus*) and non-human primates (mostly *Macaca mulatta*[1, 3–5]). It is associated with higher-order brain functions such as sensory perception, spatial reasoning, cognition, and language. Moreover, it can be further divided into frontal, parietal, occipital, and temporal lobes with distinct functional areas characterized by their neuronal connectivity. To study the association of these functional areas and the structure of the cerebral cortex, it's important to characterize the subdivisions of the cortical sheet and map each subdivision to related brain functions.

Functional areas and Brodmann areas

To this end, German anatomist Korbinian Brodmann pioneered the work over a century ago and defined over 40 Brodmann areas based on the cytoarchitecture of neurons in different subdivisions of cerebral cortex[6]. Although these areas defined by cytoarchitecture were refined with more advanced cellular and molecular techniques, nowadays, these Brodmann areas are still widely used in neuroscience studies. On the other hand, people used functional magnetic resonance imaging (fMRI) to investigate the large-scale functional network of human brain. Previous studies have implemented clustering-based methods to divide cortical vertices into a few functional networks composed of 264 functional areas based on resting-state functional connectivity (RSFC) MRI in adult brains[7, 8]. Even these cortical areas are defined with different methods, some Brodmann areas are closely related to various functionally discrete areas. For example, Brodmann area 4 is linked to primary motor cortex (M1C) while area 17 is related to primary visual cortex (V1C). To further study how the cortical areas defines the related brain functions, it's important to investigate development and arealization of the cerebral cortex.

Development of human cerebral cortex

Neural stem cells in neuroepithelium

Prior to the neurogenesis, neuroepithelial cells at the ventricular zone (VZ) of the neural tube undergo symmetric cell division and produce ventricular radial glial (vRG) cells, a predominant founding neural stem cell (NSC) in VZ. In response to the morphogen gradients, including fibroblast growth factor (FGF), Wnts, Sonic hedgehog (Shh) and bone morphogenetic proteins (BMPs), vRG cells start expressing patterning transcription factors such as EMX1, EMX2, PAX6, LHX2, and COUP-TF1[9–12]. These observations of pre-patterned expression of these transcription factors inducing region-specific neurodevelopment at early stage. Around post-conception week (pcw) 12 of human development, outer radial glia (oRG) cells are produced by asymmetric

cell division from vRG and form a distinct anatomical zone called outer subventricular zone (OSVZ)[13]. The radial glia fibers of oRG cells can extend to the pial surface, which can act as scaffolds to support the migration of subsequently developed neurons to produce supragranular layer[14]. It was reported that oRG cells may play critical roles in gyrification of cerebral cortex and the formation of cytoarchitectonics in different brain regions[15, 16]. The other daughter cell asymmetrically dividing from vRG, called truncated radial glia (tRG), emerges around pcw 14-14.5. The tRG cells stay in VZ and regrow their radial fibers. However, the radial fiber growth of tRG is confined and cannot reach the pial surface, disabling the capacity of tRG to support the neuronal migration to the cortical plate (CP)[14, 17]. It was observed in primates that neurons derived from the same group of progenitors radially migrated and resided in nearby positions of the adult cerebral cortex[18]. Hence, Rakic proposed the radial unit hypothesis, suggesting that the cortical expansion could result from the changes of the proliferative kinetics of founder cells in VZ that increase the number of radial columnar units without altering the number of neurons within each unit.[19]. Therefore, the expansive divisions of vRG and the number of oRG cells seem to be related to the tangential expansion of the cortical surface. Nonetheless, if oRG is directly linked to tangential cortical expansion is still elusive. In addition to these RG cells, intermediate progenitor cells (IPCs) exist in subventricular zone (SVZ) could divide symmetrically and differentiate into postmitotic neurons[20]. The IPCs could be more heterogeneous in terms of their morphologies, molecular properties, proliferative capacity. For example, human IPCs in OSVZ were reported to have higher proliferative capacity and could undergo expansive division[15]. Also, it was discovered that IPCs could have potential to differentiate into RG cells[21]. These studies imply that IPCs play critical roles in cell fate decision in developing human cortex. However, if IPCs in different cortical areas have different molecular properties that give rise to the area-specific phenotypes in human cortex still remains to be investigated.

Glutamatergic excitatory neurons and laminar layer formation

In the human cortex, neuronal production starts around pcw 6 and lasts through mid-gestation and possibly third trimester[22]. Two major neuronal cell classes with distinct origins are generated. Glutamatergic/excitatory pyramidal neurons and GABAergic/inhibitory interneurons are generated from RGs or IPCs in dorsal and ventral embryonic telencephalon, respectively. Also, the excitatory and inhibitory neurons constitute 70-80% and 20-30% cortical neurons, respectively[23]. Notably, two pioneer neurons, Cajal-Retzius neurons and subplate neurons, arrive at the CP even earlier than excitatory neurons and populate the marginal zone (MZ) and subplate region of the cerebral cortex, respectively. The Cajal-Retzius neurons are derived from developing pallium and play roles in coordinating the radial migration of excitatory pyramidal neurons[24, 25]. Subplate neurons, on the other hand, are derived from the progenitor cells in VZ and act as an organizer in cortical development[26]. In contrast to mouse whose Cajal-Retzius and subplate neurons mostly disappear in early development, Cajal-Retzius neurons stay in the cortical layer 1 and subplate neurons give rise to the interstitial neurons in the white matter region of human cerebral cortex. The newborn excitatory neurons are directly produced in the cortex and migrate to the layer 2-6 of CP via the radial fibers of oRGs in an "inside out" manner[27]. That is, the early-born neurons migrate to the deep layers of the cortex while the late-born neurons migrate to the upper layers. These cortical excitatory neurons combined with the Cajal-Retzius and subplate neurons overall establish the 6 laminar layers and white matter (WM) zone of the cerebral cortex. Nonetheless, the maturation of excitatory neurons continue even after their arrival to CP. The cortical excitatory neurons undergo further refinement and can be separated into different subtypes based on their laminar layer positions and neuronal projections[28]. A study performing large-scale gene profiling of human neocortex at cellular-resolution discovered species-specific molecular signatures for these excitatory neurons at different layers[29]. More recently, a study using single-nucleus RNA-sequencing (snRNA-seq) to investigate the gene expression profile of the adult human cortex at different lam-

ilar layers identified new markers of layer identity[30]. It's worth noting that these molecularly defined excitatory neurons could present in more than one laminar layer, implying the interplay of molecular identity and laminar position may come together to define the diverse cell types within and across cortical areas.

GABAergic inhibitory interneurons

GABAergic inhibitory interneurons, in contrast, are mostly generated in medial ganglionic eminence (MGE), caudal ganglionic eminence (CGE), and the pre-optic area (POA)[31–34]. The inhibitory neurons have rather diverse morphological, molecular, and connectivity properties[35]. Based on the morphological and molecular properties, MGE-derived inhibitory neurons can be further divided into PVALB, Chandelier, SST, SST CHODL subtypes while CGE-derived inhibitory neurons have LAMP5, LAMP5 LHX6, VIP, SNCG, and PAX6 subtypes. They migrate tangentially with complicated routes to CP and integrate with the cortical excitatory neurons to form local microcircuits in developing cortex[36]. Some studies have shown that the extrinsic cues from cortical neurons can guide the migration and wiring of interneurons to form local microcircuits. For example, it was shown that induction of specific types of cortical excitatory neurons in the subcortical regions cause recruitment of the layer-specific interneurons, implying that extrinsic cues from excitatory neurons can mediate the migration and positioning of interneurons[37]. Recent findings also suggest the establishment of synaptic connections between interneurons and excitatory neurons could happen in a cell type specific manner in prefrontal cortex (PFC)[38]. Therefore, the extrinsic cues and the cell type specific synaptic connections may come together to refine the integration of interneurons into the neuronal circuits.

Arealization of cerebral cortex

Intrinsic gene expression and extrinsic signaling

It's been discovered that adult human cerebral cortex has tremendous regional variations regarding the cell number, density, lamination of distinct cell types, and

specialized functions[6, 39, 40]. Over the past decades, how the functionally discrete areas form from a relatively uniform neuroepithelium during brain development was heavily investigated and people came up with two main hypotheses. The first hypothesis is called "protomap" hypothesis, stating that the pre-patterned progenitor cells are formed during neurodevelopment and these progenitor cells result in area-specific molecular features and local circuitry[41]. The second hypothesis is called "protocortex" hypothesis, suggesting that the cerebral cortex is initially homogeneous and area-specific features form in response to external stimuli, especially thalamocortical afferents [42–44]. With decades of efforts, people now realize that these two hypotheses may reconcile to explain different stages and areas of the cerebral cortex development[45, 46].

From animal models, it's been discovered that the pre-patterned neural progenitor cells form in response to the extrinsic morphogen gradients, including BMPs, Wnts, Shh, and FGF[11, 46]. In a recent study, many differentially expressed genes were found in various types of cells, including neural progenitor cells, in distinct cortical areas of early developing human brain in the second trimester[47]. Besides, they also discovered that the areal difference of the molecular signatures are rather dynamic. Specifically, they analyzed the differentially expressed genes for cell types in the excitatory lineage (RG, IPC, excitatory neurons) across human cortical areas, including PFC, motor cortex, somatosensory cortex, parietal cortex, temporal cortex, and V1C in the second trimester. Surprisingly, more differentially expressed genes were found in RGs than in excitatory neurons. In contrast, the specificity of areal markers was higher in excitatory neurons than in RGs, consistent with previous findings[44]. In line with a previous study reporting PFC and V1C neurons showed drastically different gene expression profile right after their birth[48], this recent study further discovered the RGs in PFC and V1C are already mutually exclusive, indicating two distinct gene expression programs mediating the arealization of these two cortical areas in early developing human cortex. These findings suggest that the cell-intrinsic markers of the cortical areas are already established in the neural progenitor cells (i.e.

RG), supporting the longstanding protomap hypothesis.

In addition to extrinsic morphogenic cues and the intrinsic gene expression of various types of cells during early cortical development, it's been proved that extrinsic signals from thalamus also play important roles in the neuron maturation and cortical area refinement. For example, thalamocortical stimuli can instruct the integration of MGE-derived interneurons in the mouse somatosensory cortex during early development[49, 50]. In addition to the sensory areas, it's revealed that thalamus inputs are also critical for the radial positioning and maturation of MGE-derived interneurons in the postnatal mouse PFC[51]. Besides, it's reported that thalamocortical axon input is required to drive further differentiation of the large occipital visual cortical field into primary visual area (V1C) and higher-order visual areas[52]. This finding requires the reconciliation of the prevailing protomap hypothesis with the protocortex hypothesis and implies distinct stages of arealization of visual cortex. In early stage, intrinsic genetic mechanisms drive the specification of the occipital visual cortical field. In later stages, the occipital visual cortical field with relatively uniform gene expression profile requires external input from thalamocortical axon to further differentiate into V1C and higher-order visual areas. Together, the arealization of cerebral cortex requires molecular mechanisms from both intrinsic gene expression and extrinsic signaling.

Molecularly defined cortical areas by gene expression profiles

To decipher the molecular phenotypes underlying the functionally distinct cortical areas, people have launched multiple waves of large-scale transcriptomic studies since the invention of microarray. First, Šestan's group used microarray to measure exon-level transcriptome of bulk tissues dissected from 11 topographically defined cortical areas of postmortem human brains spanning from early fetal development through old age (pcw 10 to 82 years old)[53, 54]. In their analysis, they discovered the areal transcriptomic difference follows an hourglass pattern in time, separating the human cortical development into three major stages, prenatal development, preadolescence, and adolescence. During prenatal development, they found the highest amount of differential expressed genes across areas and gradient-like expression patterns. In

the preadolescent stage, less transcriptomic difference across areas and increased synchronization of areal transcriptomes were observed. From the adolescence onward, the number of differential expressed genes across areas slightly increased but the gradient-like spatial patterns were lost, implying more subtle transcriptomic variations in the adult human cortex. To build a more comprehensive transcriptional atlas of the adult human brain (Allen Human Brain Atlas), scientists in Allen Institute of Brain Science (AIBS) performed an extensive microarray profiling of around 900 neuroanatomically precise subdivisions in two adult human brains[55]. Although the microdissection of human brain provided higher spatial resolution, they still found the neocortex displays a relatively uniform transcriptional patterns but with distinct features associated with primary sensorimotor cortices and with enriched frontal lobe expression. Also, they noticed the closer two cortical areas, the more similar their transcriptomes.

Afterward, the emergence of bulk RNA-sequencing (RNA-seq) allows people to re-evaluate the more subtle transcriptomic variations in adult human cortex with higher sensitivity and specificity. A research group leveraged bulk RNA-seq to measure poly(A)⁺ RNA fraction in 8 brain regions, including 5 cortical areas, across human, chimpanzees, gorillas, a gibbon, and macaques[56]. Surprisingly, such bulk RNA-seq profiles didn't find pronounced transcriptomic difference between cortical areas. Nonetheless, bulk RNA-seq averages the gene expression across the entire cell populations in the tissues. Thus, the areal transcriptomic variations across cortical areas are obscured, especially when the major cell types such as glutamatergic and GABAergic neurons are present in most cortical areas. With the advent of single-cell/nucleus RNA-sequencing (sc/snRNA-seq), people are able to measure the transcriptomic profiles of individual single-cells in developing and adult brains across multiple species[47, 48, 57–59]. Generally, single-cell transcriptomic analysis of developing and adult brain recapitulates the temporal hourglass patterns of areal transcriptomic divergences and also uncovers unique gene expression profiles within and between cortical areas that were previously missed by microarray or bulk RNA-seq. For example, the subtle transcriptomic differences between excitatory and inhibitory neurons in adult mouse

neocortex were disentangled. Also, an atlas work revealed that the coarsely defined subclasses of excitatory neurons were shared between most cortical areas while the inhibitory neuron transcriptomes are mostly homogeneous across the cortex[59]. Besides, the subclasses of excitatory neurons exhibited subtle different transcriptional profiles that give rise to the gradual transcriptomic shifts across the anterior-posterior and mediolateral axis in the adult mouse cortex. At the molecular level, they also discovered graded expression of some genes that have similar expression level in neighboring cortical areas and become gradually different with the increase of spatial distance in both developing and adult cortex[47, 59]. For example, the expression of marker gene *NR2F1* is expressed at higher level in posterior most V1C than anteriorly located PFC. More examples of these marker genes showing gradient patterns were reviewed in Cadwell C. R. et al.’s paper[46]. In addition to the graded expression of these genes across cortical areas, they also discovered the gradual transcriptomic differences across laminar layers in each cortical areas. For example, the laminar layer marker gene *RORB* spans across multiple layers but enriched in layer 4 of many cortical areas. Nonetheless, discrete area-specific expression of individual genes were also found in the developing human cortex. For example, the subcortical projection marker *BCL11B* and the intracortical projection marker *SATB2* co-expressed in PFC neurons but did not colocalize in V1C neurons[48]. To sum up, the gene expression profiles across cortical areas become increasingly similar during cortical development but the subtle areal transcriptomic differences in the adult brain sufficiently give rise to the functionally distinct cortical areas.

Transcriptomic cytoarchitecture of human cerebral cortex

Diverse cell types in human brain

The adult human cerebral cortex is remarkably complex, with distinct functional areas having cytoarchitectures that are made up of diverse neurons as well as

glial cells. Hence, understanding these neurons and glial cells and their spatial organization is critical to decipher their relation to complicated brain functions. However, the neurons and glial cells are rather complicated and heterogeneous. They display diverse molecular, physiological, morphological, and functional properties[60]. Thus, grouping these cells into cell types, cells with similar properties, is an important step.

To categorize cell types in the brain, people use different modalities to characterize the features of neurons and glial cells. Over a century ago, Ramon y Cajal has pioneered the work and reported stereotyped neuron types based on their shapes visualized by Golgi stain[61]. Also, people use *in vivo* electrode recording and functional imaging to investigate electrophysiology of these cells. Moreover, electron microscopy was used to study the detailed morphology and connectivity of neurons at super high spatial resolution. However, these modalities either better examine cells *in vivo* or have low throughput, making it challenging to study diverse cell types of the brain in an unbiased and comprehensive way. A quantitative, scalable, and standardizable approach is required to obtain more rigorous cell type categorization.

Molecular cell-typing by single-cell/nucleus transcriptomics

With the advent of next-generation sequencing (NGS), sc/snRNA-seq is now widely used to categorize cell types in different species and brain regions due to its high throughput and high dimensionality. It's possible to measure the whole transcriptome of thousands of cells or nuclei in one sc/snRNA-seq assay. Combined with data-driven computational clustering, cell types can be defined in a more rigorous and quantitative manner. With this scalable and quantitative technique, Zeng's group identified 133 transcriptomic cell types in mouse V1C and anterior lateral motor cortex[62]. Also, they found nearly all types of GABAergic neurons were shared in these two functionally distinct cortical areas. In contrast, most types of the glutamatergic neurons appeared in one of the two regions. In 2017, the Brain Initiative Cell Census Network (BICCN) launched a collaborative endeavor aiming to generate comprehensive brain cell atlases of mouse, monkey, and human brain[63]. In 2021, BICCN published a multimodal cell census and atlas of the M1C in mouse, marmoset and

human[64]. Integrated analysis of snRNA-seq datasets across species was performed and a consensus taxonomy of transcriptomic cell types with their hierarchical organization was reported. Specifically, 45 conserved subclasses spanning three major cell classes, including 24 GABAergic, 13 glutamateric, and 8 non-neuronal subclasses, were identified. With finer clustering, more transcriptomic cell types under each subclass were discovered. Strikingly, it was reported that human M1C alone contained 127 cell types, reflecting the highly diverse nature of the human brain cells.

From single-cell transcriptomics to spatial transcriptomics

In addition to single-cell transcriptomics, analyzing gene expression of single-cells in its spatial context is critical to link gene expression profile to anatomical, functional, and physiological properties in biological studies[65–68]. Many studies have identified more than 2000 mRNAs with different subcellular localization patterns in the somatic, dendritic, and axonal regions of neurons[69–71], indicating important roles of the spatial transcriptome in neuronal functions. For example, it was found that the 3'UTR of CAMK2A mRNA regulates dendritic mRNA localization and promotes long-term potentiation and memory formation[72, 73]. Also, some mRNAs are positioned near synapses and translated locally to maintain synaptic plasticity[74]. Although sc/snRNA-seq is powerful at profiling the whole transcriptome, the spatial information of each single-cell is lost during cell or nucleus isolation. Thus, the tool to measure gene expression of each single-cell in space is much needed.

Technologies to measure gene expression in space

Gold standard method - single-molecule fluorescence in situ hybridization (smFISH)

To measure gene expression of each single-cell in situ, Arjun et al. developed single-molecule fluorescence in situ hybridization (FISH), or smFISH. Briefly, RNA molecules in the samples were hybridized with 48 fluorophore-labeled probes to generate diffraction-limited spots visualized by fluorescence microscopy. Combined with

cell/nucleus staining, RNA spots could be assigned to cell/nucleus in proximity to profile the gene expression of each single-cell in space. Since smFISH has higher sensitivity than scRNA-seq (>95% detection rate)[76], smFISH becomes the gold standard method in spatial transcriptomics field. However, the multiplexity of smFISH is limited due to the spectral overlapping of fluorophores. Traditional wide-field fluorescence microscopes usually only image three to four fluorophores while a previously published method with special filter set-up was able to image seven fluorophores at maximum without signal cross-talk between fluorescent channels[77], restricting smFISH to measure only a few RNA species in situ. To increase the multiplexity, Linnarsson’s group further developed cyclic-ouroborous smFISH method (osmFISH)[78] which linearly scaled up the number of targeted genes by multiple rounds of FISH. They used osmFISH to measure 33 genes in 4,839 cells and reveal the spatial organization of 31 cell types in mouse somatosensory cortex. Nonetheless, higher multiplexity to measure hundreds of genes is still required to resolve more cell types or biological pathways.

Sequencing-based methods

With NGS, more sequencing-based methods such as Spatial Transcriptomics[79] and Slide-seq[80] has been developed. Generally, these methods transfer mRNA molecules in tissue sections to a surface coated with spatially barcoded DNA oligos so that the spatial information of these RNA molecules could be registered and sequenced. In detail, Spatial Transcriptomics immobilized arrayed DNA oligos to generate spatially-barcoded spots at a diameter of 100 μm and a center-to-center distance of 200 μm on a glass slide. This method was further improved with 10x Genomics barcoding technology and commercialized as 10x Visium platform with 55- μm spatially barcoded spots in late 2019. On the other hand, Slide-seq deposited 10- μm DNA-barcoded beads on a glass surface, forming a monolayer (they termed a ”puck”) to capture RNA molecules in tissue sections. Although these methods can reach transcriptome-wide gene expression of tissue sections, the spatial resolution of these methods are generally not at single-cell level. Also, the material transfer and the low capture efficiency restrict their sensitivity. It was reported the sensitivity of Spatial

Transcriptomics and Slide-seq was 6.9%[79] and 1%[80] of smFISH, respectively.

More recently, Fan’s group leveraged microfluidics and developed DBiT-seq to achieve spatial barcoding in situ[81]. Briefly, they used parallel microfluidic channels to deliver the first oligo-dT-tagged DNA barcodes with a ligation linker to tissue sections. Then, these barcodes will hybridize with mRNAs and be ligated with the second DNA barcodes delivered by microfluidic channels from a perpendicular direction, generating 2D mosaic tissue pixels. After in situ ligation and reverse transcription, the spatial information of mRNAs can be registered and sequenced. With all chemistry performed in situ, DBiT-seq could have higher capture efficiency. The sensitivity of DBiT-seq was estimated to detect an average of 15.5% of total mRNA transcripts defined by smFISH. However, the spatial resolution of DBiT-seq is limited by the 10- μ m width of the parallel microfluidic channels. So, the spatial resolution and the sensitivity still have room for further improvement.

Imaging-based methods

In addition to previous methods that relies on NGS, Je Hyuk Lee et al. developed FISSEQ to sequence crosslinked cDNA amplicons in situ on a confocal microscope[82, 83]. Generally, FISSEQ starts with reverse transcription and amplifies cDNAs by rolling circle amplification (RCA). Then, the cDNA amplicons are crosslinked in cells or tissues and sequenced using a sequencing-by-ligation strategy. Thus, FISSEQ enables transcriptome-wide profiling of gene expression in situ and preserves the cytoarchitecture of the tissue samples. Also, the RCA reaction in FISSEQ creates hundreds of sequencing templates during confocal imaging, making it easily overcome the autofluorescence background in the human tissues. Although FISSEQ substantially revolutionizes the technology by making RNA-seq in situ, it still has some disadvantages. First, most reads in FISSEQ samples are mapped to ribosomal RNAs (rRNAs) or housekeeping genes, which are not illuminating for annotating diverse cell types in complexed tissues. For example, it was reported that 82.7% reads in cultured fibroblasts measured by FISSEQ were ribosomal RNAs (rRNAs)[82]. Second, the sequencing-by-ligation procedures in FISSEQ is time-consuming and costly, making

it hard to handle large tissue samples. Therefore, a highly multiplexed and targeted method is desired to measure informative RNA species in situ.

To measure targeted genes and improve the multiplexity, many research groups developed imaging-based methods such as MERFISH[84], seqFISH+[85], STARmap[86], HybISS[87], and BOLORAMIS[88], to scale up the multiplexity exponentially by combinatorial barcoding. Take one of the first published method, MERFISH, as an example. MERFISH is generally similar to smFISH. It uses primary probes to tile RNA molecules in tissue sections and the gene-unique barcodes are molecularly realized by adding a combination of readout sequences to the sides of the primary probes. After multiple rounds of FISH, the identities of the RNA spots could be decoded by checking the gene-barcode codebook. Although each round of FISH may introduce non-specific signals, enforcing the barcodes having hamming distance (HD) of 4 and 2 could allow error-correction and error-detection capacity of MERFISH, respectively, to remove these noises. With the combinatorial barcoding, the number of detected genes increase exponentially with the number of FISH round. Since the chemistry of MERFISH is similar to smFISH, MERFISH also have high sensitivity and the spatial resolution is at single-cell level. However, the MERFISH signals are generally at low intensity in the raw images, making it hard to apply to human tissues containing high autofluorescence signals derived from lipofuscin granules[89, 90], collagen, elastin[91] and mitochondria[92, 93]. Also, it uses a high magnification objective with high numerical aperture (NA) to visualize the diffraction-limited spots, making it time-consuming when handling large tissue sample. In addition, the signal generation strategy requires probe tiling of sufficiently long RNA molecules, restricting it to detect RNA species shorter than 1.5k nt. Hence, further development is required to overcome these disadvantages.

To address these problems, Wang’s group used SNAIL probes in STARmap[86] while Daniel Gyllborg et al. and Songlei Liu et al. used padlock probes in HybISS[87] and BOLORAMIS[88], respectively, to incorporate combinatorial barcoding and signal amplification in their methods. In STARmap, a SNAIL probe is composed

of a pair of primer and padlock probes. Gene-unique barcodes are placed on the backbone of the padlock probes. The padlock probes will be circularized and amplified by RCA reaction only when the primer and padlock probe capture the targeted RNA molecules. This SNAIL design can exclude noises arise from the nonspecific hybridization. Additionally, the resulting RCA colonies or "rolonies" contain hundreds of copies of gene-unique barcodes in the form of nanoballs, fulfilling signal amplification in STARmap. Afterward, the amplified gene-unique barcodes are decoded by in situ sequencing by ligation with two-base encoding for error correction (SEDAL) strategy. This readout strategy includes extra error reduction capacity but the T4 DNA ligase reaction takes 3 hours in each readout cycle, making the whole image acquisition rather time-consuming. Similarly, in HybISS, people used padlock probes containing gene-unique barcodes on their backbone to capture cDNAs and generate rolonies in situ. Then, they used bridge probes hybridized with the barcode sequences on the rolonies and multiple rounds of FISH to decode the rolonies. This sequential hybridization readout strategy includes cycles of stripping, probe hybridization, and imaging at room temperature, enabling a more efficient image acquisition workflow. They demonstrated measurement of 119 genes in the whole coronal mouse brain section (60 mm^2) in 5 cycles of imaging and each imaging cycle took less than 5 hours with 40x objective. Although the padlock probe capture at high temperature ensure high specificity, the sensitivity of HybISS is very low. Thus, a scalable method with higher sensitivity and lower cost is much needed.

Outline of the first part of the dissertation

This work aims to develop a spatial transcriptomic tool, overcoming the limitations learned from the published methods. First, it needs to be highly multiplexed. It should be able to detect hundreds of genes in one assay to uncover diverse cell types in the human brain. Second, the spatial resolution should be at single-cell level. Third, it should be high throughput, which needs fast imaging protocol to measure more cells in large tissue samples. Most importantly, it needs to be appli-

cable to human tissues, especially human brain tissue that has high autofluorescence background from lipofuscin granules. To this end, we developed Decoded Amplified targeted Transcripts with Fluorescent in situ Hybridization or DART-FISH. The first part of this dissertation demonstrates the framework of DART-FISH, benchmarking and validation using human M1C in Chapter 2. During DART-FISH development, I collaborated with my lab member, Kian Kalhor and Mahsa Nafisi. We applied this tool to study different human tissues. Kian and Mahsa applied DART-FISH to study human kidney and human lung tissue, respectively, while I applied DART-FISH to uncover the cytoarchitecture of neuronal and non-neuronal cells in human cerebral cortex.

Outline of the second part of the dissertation

In addition to study the human M1C, I am interested in the transcriptomic variation and cellular heterogeneity of the cortical sheet in the adult human brain. Besides, investigating the gene regulatory network maintaining the cell identity and spatially variable genes is also critical. To this end, the second part of this dissertation details a collaborative work between Kun Zhang’s lab, Jerold Chun’s lab and Bing Ren’s lab. Jerold’s lab collected adult human cerebral cortex tissue blocks, including M1C, V1C, primary somatosensory cortex (S1C), primary auditory cortex (A1C), middle temporal gyrus (MTG), dorsolateral prefrontal cortex (DFC), anterior cingulate cortex (ACC), angular gyrus (AnG), and frontoinsular cortex (FI). Dr. Carter Palmer from Chun’s lab provided the human brain sections and dissociated nuclei for the downstream molecular assays. I applied DART-FISH to spatially profile the expression of 484 genes in these brain sections while a postdoctoral fellow, Dr. Jinghui Song, used SNARE-seq2 to profile the whole transcriptome and chromatin accessibility of individual dissociated nuclei. Dr. Nathan Zemke from Ren’s lab led his team to measure the transcriptome and chromatin accessibility of dissociated nuclei by commercially available 10x multiome assay for cross-validation.

With wide sampling of human cerebral cortex, the variations of gene expres-

sion and the organization of cell types along the rostrocaudal axis of human cortex are revealed by DART-FISH. In addition, SNARE-seq2, a dual-omic method to profile the transcriptome and chromatin accessibility information of individual single-nuclei, were used to investigate the gene regulatory network[94]. SNARE-seq2 is a single-cell sequencing technology to profile the transcriptome and chromatin accessibility of individual single-cells. However, SNARE-seq2 assays were carried out in dissociated nuclei. Again, the spatial information is lost. Thus, the gene expression profile of DART-FISH cells are leveraged to integrate the pre-annotated SNARE-seq2 data in space. Ultimately, we expected to generate spatially resolved cell atlases, including the transcriptome and chromatin accessibility information, of adult human cerebral cortex. My involvement in this collaborative work includes DART-FISH data generation and integrative analysis between DART-FISH and SNARE-seq2 datasets. At the moment I write this dissertation, the comprehensive analysis of chromatin accessibility is still ongoing. So, I only include the results of aligning the DART-FISH cells with SNARE-seq2 nuclei to analyze the transcriptomic divergence across the cortical areas in Chapter 3.

CHAPTER 2: DEVELOPMENT OF DECODING AMPLIFIED TARGETED TRANSCRIPTS WITH FLUORESCENCE IN SITU HYBRIDIZATION(DART-FISH)

Introduction

Single-cell gene expression in its spatial context plays a critical role in the development of complexed cellular organization in multicellular organisms[65, 68]. Analysis of RNAs in heterogeneous tissues is especially important in structure-focused fields such as embryo development, histopathology, and neuroscience. The development of smFISH allowed us to simultaneously measure a few RNA species in their spatial contexts by visualizing fluorophore-conjugated probes tiling the RNA molecules[75, 95]. Due to its high sensitivity, smFISH has become the gold standard of spatial transcriptomics. People used it to prove the important roles of RNA localization in cell migration, neuron connectivity, and local protein synthesis[67, 96]. Nonetheless, the spectral overlapping of fluorophores limits its multiplexity.

With the advent of next-generation sequencing (NGS), a few sequencing-based spatial transcriptomic methods were developed[79–81, 97–99]. In principles, these methods transfer RNA molecules in tissue sections to a surface coated with spatially barcoded oligos. After library preparation, the spatial information of RNAs is registered and sequenced. Although these methods expand the multiplexity to the transcriptome-wide, they have low capture efficiency, limiting their sensitivity. Besides, the spatial resolution of these methods rely on fabrication of the spatial features on the surface and lateral diffusion, which prevent these methods from obtaining single-cell level spatial resolution[100].

Recently, combinatorial barcoding was used to exponentially increase the mul-

multiplexity of many imaging-based methods. In general, they used different configurations of probes to capture the targets and the barcodes were placed on the backbones of the probes, which can be decoded with multiple rounds of FISH[84, 85]. Although these methods have exponentially increased the multiplexity by 2-3 orders of magnitude within tens of rounds of FISH, they need their probes to tile sufficiently long transcripts ($> 1.5\text{kb}$) to generate detectable signals, restricting us from analyzing short but important molecules such as neuropeptides and interferons[84, 101]. Besides, the low SNR of the detected transcripts enforced these methods to use high magnification objective with high numerical aperture (NA) during image acquisition, making them time-consuming to image large ROIs. Also, the low SNR limits their application to human tissues, which has high autofluorescence signals from the lipofuscin granules[89, 90], proteins such as collagen and elastin[91], and mitochondria[92, 93]. Therefore, a single-cell level, highly multiplexed and high throughput spatial transcriptomic method that can be directly applied to human tissues is much needed.

In this chapter, we developed Decoded Amplified taRgeted Transcripts with Fluorescence in situ Hybridization or DART-FISH (DART-FISH), a highly multiplexed RNA imaging method to measure spatial distribution of hundreds of RNA species in large tissue sections at single-cell level. The key technical features include a robust barcoding scheme, a cost-effective padlock probe production protocol, in situ padlock probe capture and amplification, a cytoplasmic stain called RiboSoma, fast decoding, and a computational framework called SparseDeconvolution (SpD) Decoding for decoding features at the pixel level from optical crowded regions of fluorescent images. We benchmarked the DART-FISH method by measuring 121 RNA species in large human M1C sections ($\sim 30\text{ mm}^2$). In addition, we validated its sensitivity and specificity by comparing it to RNAscope, a commercially available smFISH method. Furthermore, DART-FISH successfully recapitulated the cytoarchitectures of 20 neuronal and non-neuronal cell types, detected short neuropeptide genes (*SST* and *NPY*), and validated a new deep layer neuron marker (*TMSB10*) in human M1C.

DART-FISH workflow

Reverse transcription and gel embedding

DART-FISH involves padlock probe capture and rolling circle amplification (RCA) to detect targeted genes in situ and amplify their signals. Afterward, a highly robust decoding process including sequential isothermal hybridization is used to spatially resolve the targeted genes. (Figure ??a). In detail, fresh-frozen tissue sections were fixed paraformaldehyde (PFA) and permeabilized. The RNA molecules in the tissue sections are reverse-transcribed with a mixture of random and poly-dT primers concatenated with a 5' handle to access the retention of cDNAs in situ. This built-in cDNA labeling allows us to visualize the locations of cell bodies in the tissue sections during imaging (Figure ??b); thus, we call it RiboSoma for its usage of deoxyRibonucleic acid to label Soma of single-cells. It's worth noting that crosslinking cDNAs to the polyacrylamide (PA) gel after the reverse transcription step enhances the RiboSoma signals (Supplementary Figure ??a), implying better retention of cDNAs. With better retention of cDNAs in situ, it results in 1.5-fold increase of count per gene compared to gel formation after RCA (Supplementary Figure ??b-c). Therefore, RiboSoma acts as a good checkpoint to evaluate cDNA content in the tissue sections, providing a quality control for starting tissue sections.

Padlock probe capture and signal amplification

After cDNA crosslinking and RNA digestion, cDNAs are hybridized with a library of gene-specific padlock probes to capture targeted genes. In addition to the target sequences, the decoder sequences and the universal sequence on the padlock probes are used for combinatorial gene barcoding and RCA, respectively. These padlock probes are then circularized at high temperature to ensure specificity during padlock probe capture[102, 103]. Then, the RCA primers hybridizing to the universal sequences on circularized padlock probes and Φ 29 DNA polymerase initiate RCA, generating RCA colonies (or rolonies) in situ. These rolonies contain hundreds of

copies of decoder sequences in the form of DNA nanoballs, which lead to high signal intensity and overcome the autofluorescence in the human tissue sections during image acquisition. Next, rolonies are covalently bonded to the PA gel, securing their positions during multiple rounds of FISH and imaging. Later, the DART-FISH samples could be accessed during anchor round imaging. In anchor round imaging, anchor probes are hybridized to the universal sequences on all rolonies and the 5' handle of all cDNAs to visualize the spatial distribution of all rolonies and cell bodies in the tissue sections, respectively (Figure ??b).

High multiplexity and fast readout strategy

To obtain high multiplexity within several rounds of imaging, gene-specific barcodes are molecularly realized by placing a combination of 20-mer decoder sequences on the backbone of padlock probes. These 20-mer decoder sequences are derived from Illumina BeadArray technology and have minimal cross-hybridization[104]. During n rounds of image acquisition, each barcode is "on" in k and "off" in the other rounds, as illustrated in Figure ??c. In "on" cycle, the barcode lights up in one of the three fluorescent channels and stays dark during the "off" cycle. Physically, rolonies emits fluorescent signals when their decoder sequences are hybridized with their corresponding decoding probes in "on" cycles. After each round of imaging, the decoding probes are stripped off, preparing the sample ready for the next round of imaging[105] (Figure ??e). Thus, $\binom{n}{k}3^k$ barcodes can be exponentially generated with n rounds of imaging. As an example in Figure ??, 540 barcodes are generated with 6 rounds of imaging using 3-on-3-off encoding scheme, allowing us to measure hundreds of genes. This barcoding scheme were proved to be robust with wide applications by Illumina's gene expression, SNP genotyping, and DNA methylation arrays[104, 106, 107]. This sequential hybridization readout strategy avoids complicated temperature control setups and enzymatic reactions in sequencing-by-ligation readout strategy commonly used in FISSEQ, STARmap, and BOLORAMIS. Besides, this sequential hybridization process is fast and robust by making the short decoding probe hybridization and stripping happen at room temperature. Thus, the scalable barcoding scheme and

fast readout strategy make it highly multiplexed and feasible to handle large tissue sections.

Cost-effective padlock probe production

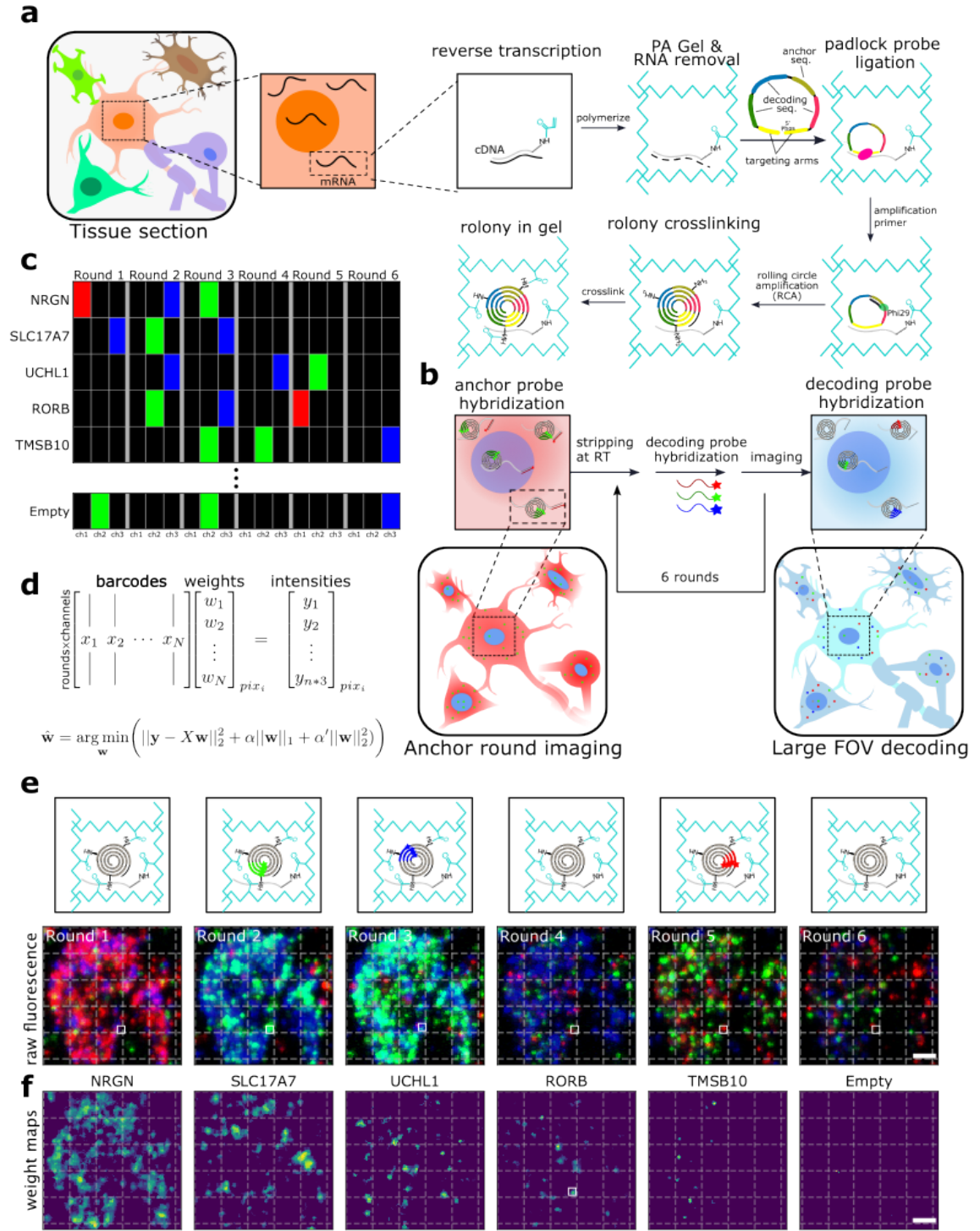
It has been proved that increase the number of padlock probes per gene can improve the sensitivity of detecting the targeted genes in situ[108]. A widely used strategy in many methods such as BARseq2[108], HybISS[87], and ExSeq[109], is to pool chemically-synthesized padlock probes together. Although this strategy is applicable to target small gene panels, it becomes tremendously pricey when thousands of padlock probes are required to target hundreds of genes. To address this practical problem, we directly leveraged an enzymatic protocol to massively produce thousands of padlock probes from a seed oligo pool synthesized on microarrays[110], which substantially reduce the cost of padlock probe production (Supplementary Figure ??c). As an example, we spent less than 25% of the direct synthesis option with our protocol when we produced 5,097 padlock probes to target 121 genes (Each gene was targeted by 50 padlock probes at maximum). It's worth noting that the cost per padlock probe could be further reduced by combining more padlock probes in the seed oligo pool. In contrast, the cost per padlock probe stay constant for the direct synthesis option. Hence, producing 200,000 padlock probes to target 400 genes is almost 10 times as expansive as the seed oligo pool synthesized on microarrays. To fully leverage this cost-effective padlock probe production strategy, we can even pool multiple probe sets targeting genes in different organs together during seed oligo pool synthesis and amplified each probe set separately. It costs us a fraction of the upfront cost of the direct synthesis option and make a big probe set affordable to any regular lab.

SparseDeconvolution (SpD) Decoding resolves optical crowding

As we increase the multiplexity and sensitivity, colonies become crowded and start to cause optical crowding, which may hinder subsequent colony decoding. To address this problem, some research groups use gel expansion strategy to physically

enlarge the tissue sections[109, 111, 112]. This strategy effectively extends the physical distance between rolonies, reducing the overcrowding of these nanoballs in the tissue sections. However, the physical enlargement of the tissue sections nullifies our efforts to shorten the imaging time by fast decoding because it takes longer time to acquire images of the same ROIs after sample expansion and thus lower the throughput[109]. Therefore, a computational solution to the roloniy decoding can not only solve the optical crowding problem but also rescue our efforts to reduce the imaging time spent on a large tissue sample. Given the physical size of the rolonies ($< 1 \mu\text{m}$ [82]) and the pixel size ($0.3 \mu\text{m}$ with 20x objective) of the confocal images, each pixel will overlap with a few rolonies at maximum. Meanwhile, a small fraction of all possible barcodes are used, making it the possible to deconvole mixtures of barcodes from the fluorescent intensities values of each pixel. Here, we developed the *SparseDeconvolution* (SpD) decoding algorithm: we formalized this deconvolution as a regularized linear regression problem, where barcodes can linearly combine with different weights to form the measured fluorescent intensities of each pixel and finding the optimal combinations of weights under a condition that promotes sparsity (Figure ??d). We solve this problem for each pixel and obtain weight maps for every single barcode/gene (Figure ??f), followed by filtering and aggregating the neighboring pixels to form spots (Supplementary Figure ??f-h). To set a quality control of this deconvolution procedure, we add empty barcodes that are not used in the gene-barcode map to the codebook. Although the fraction of empty barcode is 5-8% of the used barcodes, the fraction of spots decoded as empty is less than 0.25% (Supplementary Figure ??f-h). With this computational framework, we can decode the rolonies in the presence of optical crowding and increase the throughput by imaging the tissue sections with a 20x objective.

Figure 2.1 DART-FISH workflow: Decoding Amplified targeted Transcripts with Fluorescence in situ Hybridization (a) DART-FISH schematics. RNA molecules in the fresh-frozen and formaldehyde-fixed tissue section are primed with primers with a 5' handle sequence and reverse-transcribed. A polyacrylamide (PA) gel is cast on the tissue, incorporating the cDNA molecules. After RNA removal, padlock probes are hybridized to cDNA and circularized, followed by rolling circle amplification (RCA) to create rolonies. Rolonies are further crosslinked to the gel. (b) Imaging DART-FISH samples. Samples subsequently go through anchor round imaging followed by decoding rounds. In anchor round imaging, fluorescent probes complementary to the RCA primer, present on all rolonies, and the 5' cDNA handle, present on all cDNA molecules, are hybridized at room temperature to visualize the distribution of rolonies and the shape of the somas, respectively. After imaging, the fluorescent probes are stripped and washed away at room temperature. In the decoding rounds, round-specific decoding probes are hybridized, imaged and stripped. This procedure is repeated n times ($n=6$ in this example). (c) An example codebook for DART-FISH. Each gene is barcoded such that the corresponding rolonies show fluorescent signal in k rounds of decoding and remain off in other rounds. 5-10% of the codebook consists of empty barcodes that don't have representative padlock probes and are only used for quality checks in the decoding pipeline. (d) The intensity of pixels across n rounds of 3-channel imaging is modeled as a combination of the barcodes in the codebook. The decoding is formulated as a regularized linear regression, such that more barcodes do not participate in the combination. (e) Example of fluorescence in situ hybridization on rolonies in the gel. The lower panel shows the maximum intensity projection of the fluorescent images across 6 decoding rounds and 3 channels. (scale bar 5 μm) The upper panel is a cartoon of the decoding of a *RORB* spot corresponding to the white square. (f) Lasso maps. Each Lasso plot represents the gene weights of *NRGN*, *SLC17A7*, *UCHL1*, *RORB*, *TMSB10*, and Empty spots in (e) learned by SparseDeconvolutionDecoder (scale bar 5 μm).



Benchmarking and validation

Measurement of 121 genes in human primary motor cortex

As a proof of principle, we assessed the performance of DART-FISH by applying it to measure 121 genes in a 6.9-by-4.3 mm² postmortem human M1C brain section with 6 rounds of decoding imaging. Human M1C acts as a good standard to compare with previous studies because the anatomy, function, and gene expression of M1C were widely studied at single-cell level[64, 113–115]. Besides, the archived human brain samples have high autofluorescence[116] and generally lower RNA quality[117], creating great challenges to the imaging-based spatial transcriptomic methods.

Reproducibility and spot assignment

To measure 121 genes, including curated marker genes to uncover neuronal and non-neuronal cell subclasses in human M1C, we designed 5,097 padlock probes to target these genes (Supplementary Table 2). We assigned the gene-specific barcodes to these genes using a 3-on-3-off barcoding scheme in a codebook. It takes us approximately 10 hours to finish the anchor round imaging, 6 rounds of decoding imaging, and nuclear staining of this around 30 mm² human M1C section. Subsequently, we obtained 2,008,260 transcripts (0.2% empty rate with 8 barcodes assigned to empty) after the image preprocessing and SpD decoding. We carried out the DART-FISH assay to two parallel sections. As shown in Figure 2.2b, the expression level of these 121 genes was highly correlated (Pearson’s correlation coefficient $r^2=0.988$) between the two replicates, demonstrating high reproducibility of DART-FISH.

Next, we segmented the cell bodies using RiboSoma, which depicted the boundaries of cell bodies better than nuclear staining (Supplementary Figure S4a, b), and assign the transcripts to cell bodies in proximity (≤ 3 μ m). Other transcripts were filtered out prior to cell annotation. It’s worth noting that a higher fraction of *MBP* transcripts were localized outside the cell bodies (93%, Supplementary Figure S4d), co-localizing with RiboSoma in the extrasomatic space (Supplementary Figure S4e).

This observation reflects the local translation of *MBP* transcripts at the axon-glia contact sites[118]. Overall, we detected 802,361 transcripts assigned to 26,646 cells with an average of 30 transcripts and 11 unique genes per cell (Figure 2.2c).

Specificity of DART-FISH

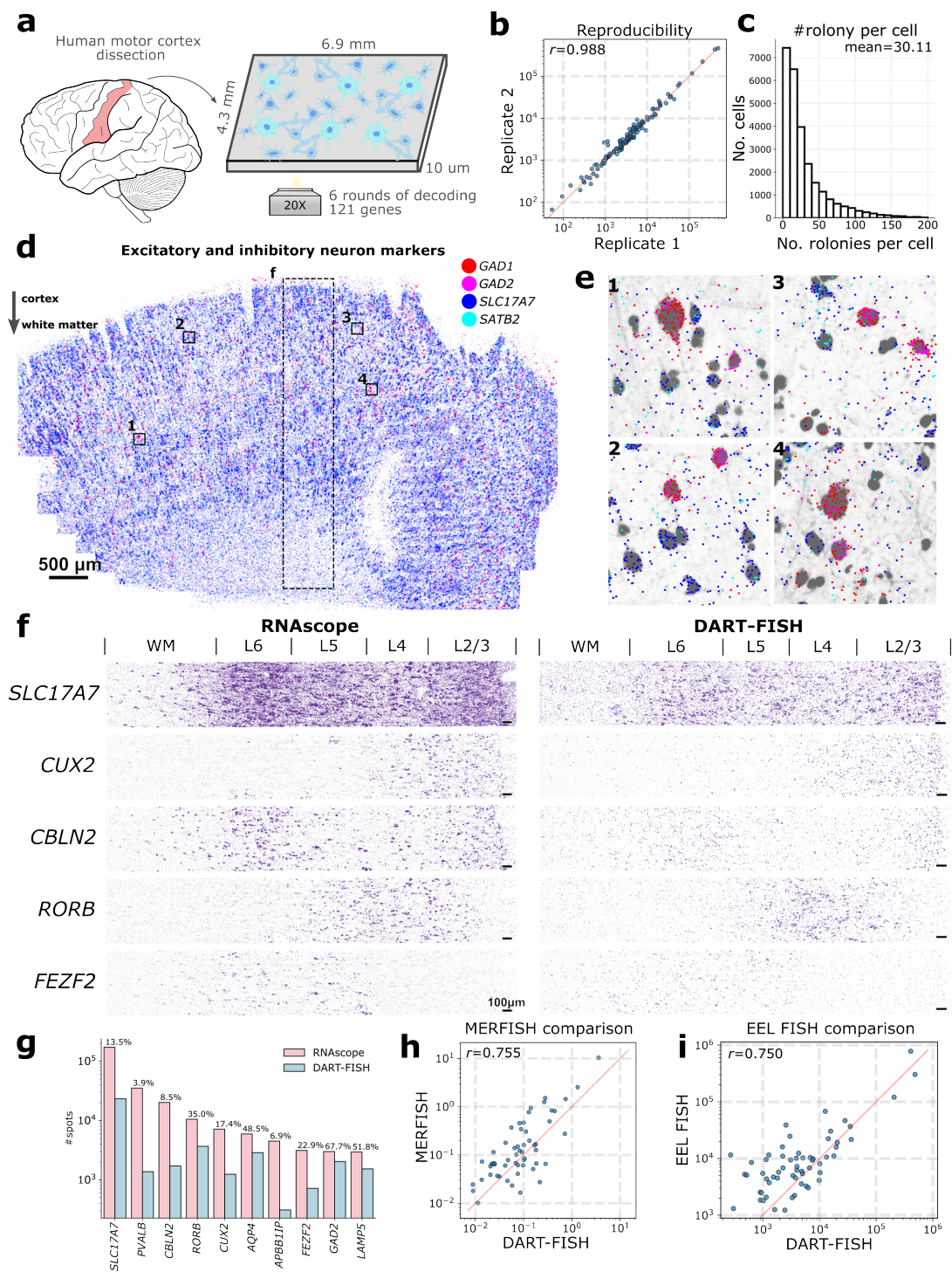
To access the specificity of DART-FISH, we first inspected the spatial localization of *SLC17A7* and *SATB2* transcripts in excitatory neurons and *GAD1* and *GAD2* in inhibitory neurons. As shown in Figure 2.2d-e, the *SLC17A7* and *SATB2* transcripts were aggregated in the soma of excitatory neurons with mutual exclusiveness to the *SLC17A7* and *SATB2* transcripts in inhibitory neurons. Second, we validated and compared the spatial localization of 10 marker genes with the results of RNAscope, a commercially available smFISH method, carried out on a parallel M1C tissue section. As shown in Figure 2.2f and Supplementary Figure S4f, the spatial localization of these marker genes in the same ROIs had high concordance between DART-FISH and RNAscope samples. Specifically, the pan-excitatory neuron marker, *SLC17A7*, had pronounced enrichment in the L2-L6 cortical layers. Also, *CUX2*, *RORB*, and *FEZF2* were enriched in supragranular, granular, and infragranular layers of the neocortex, respectively, which is consistent with previous studies[29, 119–122]. Moreover, the localization of *CBLN2* in cortical layers 2/3 and 5/6 was also in agreement with a previous study[123]. Altogether, these results imply that DART-FISH has sufficient specificity to resolve the laminar layer-specific marker genes in human M1C.

Sensitivity of DART-FISH

To access the sensitivity of DART-FISH, we compared the number of the 10 validated transcripts in a similar ROI between DART-FISH and RNAscope samples. Since RNAscope has the highest sensitivity among the spatial transcriptomic methods, we directly took the ratio of the number of transcript in DART-FISH and RNAscope samples for each gene to estimate the sensitivity of DART-FISH. We noticed that the estimated sensitivity varied across genes and ranged from 3.9% to 67.7% (Figure 2.2g). Also, we compared the DART-FISH results to MERFISH[124] and EEL

FISH[125] method that also had publicly available datasets from the human brain. As shown in Figure 2.2h and i, the Pearson’s correlation coefficients of the expression level of common genes with MERFISH and EEL FISH dataset were 0.755 and 0.750, respectively. Given the human brain regions and the sensitivity of each technology are different, we consider these results high concordance. To sum up, DART-FISH is a reproducible spatial transcriptomic method with sufficient sensitivity as well as specificity to detect hundreds of genes in their spatial context and it is able to reveal biological insights in the human brain sections despite the high autofluorescence background.

Figure 2.2 Benchmarking DART-FISH on the human motor cortex (a) Parallel sections were taken from a dissected post-mortem human M1C tissue block. Spatial distribution of 121 RNA species was measured by DART-FISH with 6 rounds of decoding imaging. (b) Scatter plot showing reproducibility between parallel tissue sections. Each dot represents the total count of each gene detected in each replicate. (c) The histogram for the number of high quality decoded rolonies per cell. (d) Spatial distribution of excitatory neuron markers (*SLC17A7* and *SATB*) and inhibitory neuron markers (*GAD1* and *GAD2*) in the whole tissue. The dashed rectangular box delineates the ROI in (f). Scale bar 500 μm . (e) Zoomed-in views to show the separation of excitatory and inhibitory markers at single-cell level in 4 ROIs delineated in (c). (f) Validation of DART-FISH by RNAscope. Spatial distribution of *SLC17A7*, *CUX2*, *CBLN2*, *RORB* and *FEZF2* across the cortical layers measured by RNAscope (left) and DART-FISH (right). Scale bar 100 μm . (g) Quantitative comparison of counts for *SLC17A7*, *PVALB*, *CBLN2*, *RORB*, *CUX2*, *AQP4*, *APBB1IP*, *FEZF2*, *GAD2*, and *LAMP5* in DART-FISH and RNAscope in a similar ROI. Percentages represent total spots detected in DART-FISH divided by total spots detected in RNAscope multiplied by 100. (h) Comparing DART-FISH and MERFISH[124] (H18.06.006.MTG.4000.expand.rep2). Each dot represents the mean count per cell for the 56 shared genes. (i) Comparing DART-FISH and EEL FISH[125] (data from visual cortex). Each dot represents the total count of the 60 shared genes.



Organization of cell types in the human primary motor cortex

Spatially resolved excitatory, inhibitory, and non-neuronal subclasses in the human primary motor cortex

To access the capability of DART-FISH to uncover the diverse cell types in human M1C, we next performed cell annotation of DART-FISH cells. We began with clustering the DART-FISH cells and match the clusters to the highest correlated subclass from a recently published human M1C snRNA-seq dataset[113, 126] (Figure 2.3a and b, Supplementary Figure S5). We uncovered 20 subclasses from the major excitatory, inhibitory, and non-neuronal cell classes which made up 24.3%, 10.6%, and 65.1% cells, respectively, in the human M1C (Figure 2.3c-g). For excitatory neurons, we successfully recapitulated the laminar layer structures with L2/3 IT subclass at the upper layer and L6b/CT subclass at the deeper layer along the cortical depth (Figure 2.3b-d). This observation is in line with the evolutionarily conserved organization of excitatory neurons in mammalian M1C[113]. Also, we noticed the L6 IT Car3 subclass was localized more superficial than the L6 IT subclass in the deep layer, consistent with the recent observations in the human primary visual cortex and middle temporal gyrus[126, 127](Figure 2.3d). In contrast, the inhibitory neurons showed wider localization along the cortical depth, likely due to their developmental origins outside neocortex[128, 129] For example, the VIP subclass was enriched in layer 2-3, in agreement with the previous studies in the mouse brain[115, 130] (Figure 2.3e). In addition, we observed some excitatory and inhibitory neurons localized in the white matter (WM) region, which are likely the remnants of the subplate neurons discovered in previous studies[131, 132]. For non-neuronal cells, we saw oligodendrocytes (Oligos) appearing at layer 4 and are enriched in the WM region[133] even though the oligodendrocyte progenitor cells (OPCs) were uniformly distributed across the tissue section[134] (Figure 2.3f).

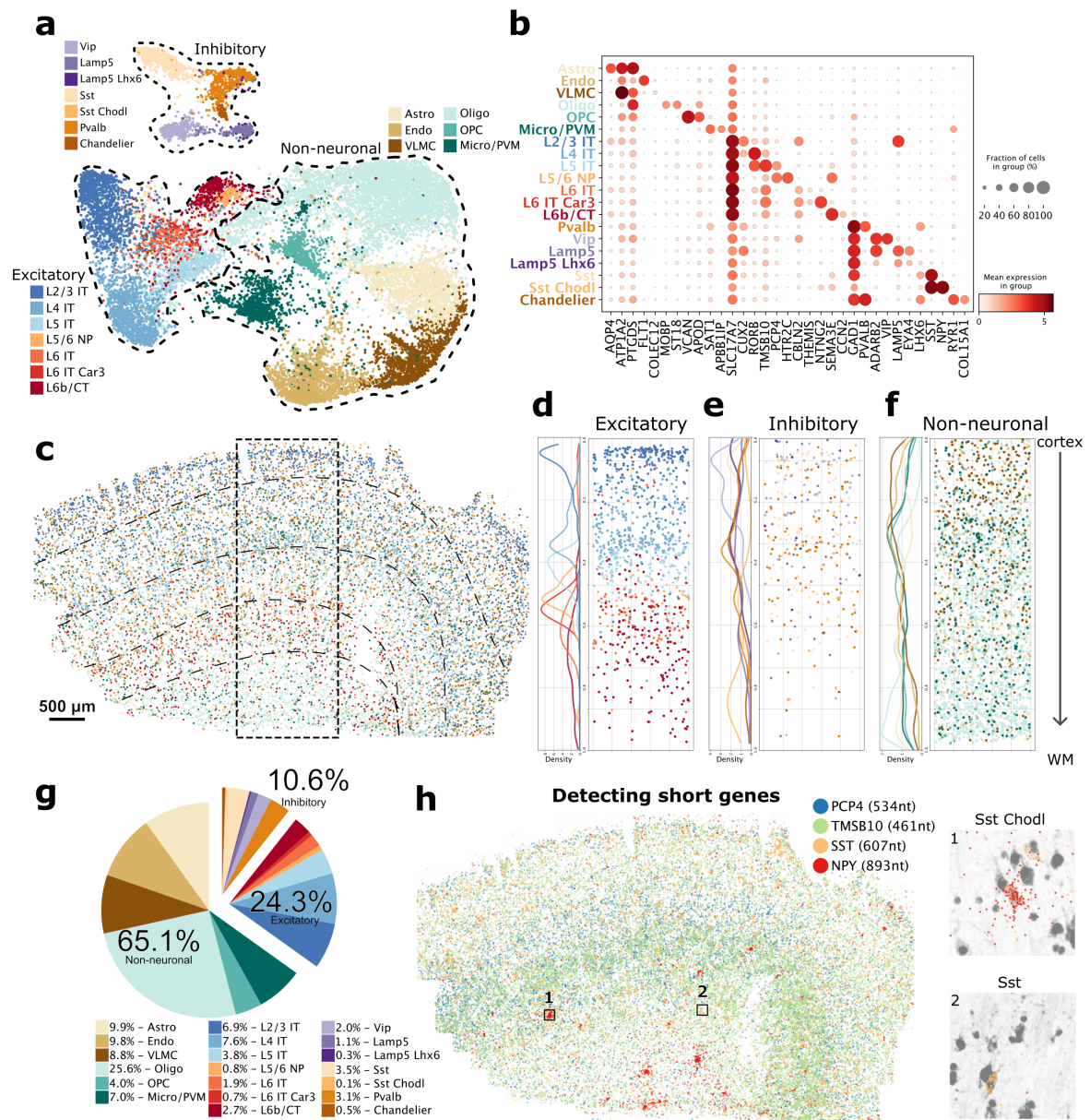
Short gene and rare cell type detection

We further assessed the capability of DART-FISH to detect short genes ($< 1.5\text{kb}$). Those smFISH-based methods (MERFISH, seqFISH+) use tens of fluorophore-tagged probes to tile sufficiently long transcripts to generate detectable fluorescent signals. In contrast, one padlock probes successfully bind to the target is sufficient to generate detectable signals in DART-FISH. Thus, DART-FISH has higher chances to detect those short RNA species. To further boost our chances to detect short transcripts, we slightly modified our probe design strategy to allow overlapping of target sequences of padlock probes targeting RNA species $< 1.5\text{kb}$ (Supplementary Figure ??b, *NPY* as an example). We selected a list of 33 genes $< 1.5\text{kb}$, including literature-curated genes and some computationally-derived genes in the brain (Supplementary Table 2). It's worth noting that we were able to uncover a rare inhibitory neuron subclass, Sst Chodl (0.1% abundance, Figure 2.3g), by targeting short neuropeptide gene SST (607 nt) and *NPY* (893 nt) (Figure 2.3b and h). Also, we observed Sst Chodl cells were mostly localized in deep layers, in line with the observation in a previous study[135]. In addition to these short neuropeptides, DART-FISH also captured pronounced localization of other two short RNA species, *PCP4* (534 nt) and *TMSB10* (461 nt) (Figure 2.3h). It was reported in the mouse cerebral cortex that *PCP4* could be a layer 5-6 marker[136]. Nonetheless, pronounced localization of *TMSB10* in the deep layer was not reported. Thus, *TMSB10* could be a novel deep cortical layer marker gene.

Subsequently, we quantify the efficacy of DART-FISH to detect the short genes. To do so, we calculated the Pearson's correlation coefficients of their average gene expression at the subclass level between DART-FISH and snRNA-seq (Supplementary Figure S5c). We found 25 of the selected 33 short genes (75%) and 81 of 88 longer genes (92%) showed higher correlation than 0.5 (Supplementary Table 2). This is comparable to a MERFISH dataset targeting 250 genes (88% with Pearson's correlation > 0.5 , Supplementary Figure S5c). Taken together, we demonstrated that DART-FISH is able to uncover the organization of the diverse cell types in the human M1C and the

rare cell types can be detected by resolving the short genes in their spatial context.

Figure 2.3 **DART-FISH mapping cell types in human M1 cortex.** (a) UMAP of all excitatory neurons (L2/3 IT, L4 IT, L5 IT, L5/6 NP, L6 IT, L6 IT Car3, and L6b/CT), inhibitory neurons (Pvalb, Vip, Lamp5, Lamp5 Lhx6, Sst, Sst Chodl, and Chandelier), and non-neuronal (Astro, Endo, VLMC, Oligo, OPC, and Micro/PVM) cell types. (b) Dot plot of marker gene expression across cell types (c) Spatial distribution of cell types in the whole M1 tissue section, using the same color palette in (a). The dashed rectangular box delineating the same ROI in (d), (e), and (f). (d), (e), (f) show the spatial distribution of excitatory neurons, inhibitory neurons, and non-neuronal cell types, respectively, across the upper cortical layer (top) and the WM zone (bottom). (g) Piecharts depicting the relative amount of both major cell classes (left) and cell subclasses (right). (h) Spatial distribution of short RNA species (≤ 1250 nt), *PCP4*, *TMSB10*, *SST*, and *NPY*, in the M1 tissue section. *PCP4* and *TMSB10* are L5 and L6 markers, respectively. *SST* and *NPY* are SST and SST Chodl cell markers, respectively. Example of SST and SST Chodl cell are shown in ROI1 and ROI2 marked by square black boxes.



Discussion

In this chapter, we described DART-FISH, a highly multiplexed and high throughput in situ RNA imaging method. Also, we showcased it’s applicable to human brain tissue sections that contain high autofluorescence signals. Moreover, we benchmarked and validated its reproducibility, sensitivity as well as specificity and used it to recover the cellular organization of 20 neuronal and non-neuronal cell types in human M1C. Specifically, DART-FISH is able to recapitulate the laminar layer structures of excitatory neurons while the inhibitory neurons generally shows wider and sparser distribution along the cortical depth. On the other hand, the non-neuronal cells are ubiquitous in the human M1C with oligodendrocytes having pronounced enrichment in the WM zone.

In addition, we provided a cost-effective padlock probe production protocol and fast decoding procedures that allow us to easily handle large tissue sections. The cost of padlock probe synthesis per gene scales sub-linearly by using our padlock probe production protocol, which enables large-scale studies that require thousands of padlock probes to target tens of thousands of genes. Besides, the rolon generation in DART-FISH allows amplification of the decoder sequences on the rolonies by over 100-fold and the increase of its physical size. Thus, the barcodes on the rolonies can be easily detected via hybridization of short fluorescent decoding probes to the decoder sequences at room temperature in 5–10 minutes. Unlike ligation-based readout strategies used in other methods (FISSEQ, STARmap, BOLORAMIS), this sequential hybridization decoding process in DART-FISH doesn’t require specialized equipments for temperature control and is relatively fast due to its isothermal stripping, washing, and hybridization procedures. Altogether, this easy and straightforward rolon generation permits the fast decoding process by using 20x objective, making it feasible to handle centimeter-sized tissue sections with a traditional confocal microscope.

Although some targeted spatial transcriptomic methods such as HyBISS and BARseq2 also use padlock probes to target cDNAs, DART-FISH differs from them

by additional chemical treatments of cDNAs. First, cDNAs in DART-FISH were covalently crosslinked to PA gel, which significantly enhances the sensitivity of DART-FISH by retaining more cDNA templates for the subsequent padlock probe capture and colony generation procedures. Second, we introduced RiboSoma to label cDNAs, which becomes a more effective marker to depict cell bodies than traditional nuclear staining. We expect this tool can be highly useful for cell body segmentation in thicker tissue sections.

Besides, the RCA-based methods generate physically large colonies, which cause optical crowding in fluorescent images when more genes are targeted with higher sensitivity. To address this problem, we developed a computational framework (SpD) that deconvolve the mixture of barcodes at each pixel based on the redundancy in the barcode space. This method improved the decoding efficiency compared to naive decoding methods directly filtering out pixels that have optical crowding issue[137]. Furthermore, sophisticated deconvolution methods that share information between neighboring pixels may further improve the decoding efficiency[138, 139]. As the field is moving towards simultaneously measuring more genes, pixel-based deconvolution decoding methods like SpD could be increasingly important.

Finally, the sensitivity of DART-FISH still have some room to improve. It is widely recognized that the efficiency of reverse transcription is capped at 20% so that some genes may not have sufficient cDNA templates for our padlock probes to capture. Thus, skipping the reverse transcription step and design padlock probes to target RNA molecules directly become a straightforward modification to improve DART-FISH sensitivity. Other than the moderate sensitivity, DART-FISH is a streamlined method that can be easily extended in many aspects. For example, antibody staining of membrane receptors or extracellular matrix proteins can be easily incorporated into the DART-FISH workflow to investigate the cell-cell interactions in healthy and pathological niches[140]. Besides, expansion microscopy and tissue clearing procedures can be combined with DART-FISH to spatially resolve subcellular localization of RNAs in finer structures of thicker tissue sections[141].

Acknowledgement for Chapter 2

I want to acknowledge all people involved in the DART-FISH tool development. The prototype of DART-FISH was built by Ho Suk Lee and Matthew Cai using cell lines and mouse brain tissue sections. Kian Kalhor, Mahsa Nifisi, and me optimized the DART-FISH using human tissues, including human kidney, lung, and brain. I particularly tailored the DART-FISH protocol to human brain tissue sections and performed RNAscope experiments for benchmarking and validation. Dr. Carter Palmer from the Chun lab provided the human brain tissue sections. The SparseDeconvolution Decoding (SpD) algorithm was fully derived from Kian Kalhor's computational expertise. The DART-FISH work on human brain in Chapter 2 were merged with DART-FISH work on human kidney from Kian Kalhor and published on bioRxiv. I am the co-first author of that paper and the primary author of this chapter.

Chapter 2 is based on the following publication:

Kalhor, Kian*; Chen, Chien-Ju*; Lee, Ho Suk; Cai, Matthew; Nafisi, Mahsa; Que, Richard; Palmer, Carter; Yuan, Yixu; Zhang, Yida; Song, Jinghui; Knoten, Amanda; Lake, Blue B.; Gaut P. Joseph; Keene, Dirk; Lein, Ed; Kharchenko, Peter V.; Chun, Jerold; Jain, Sanjay; Fan, Jian-Bing; and Zhang, Kun. 2023. Mapping Human Tissues with Highly Multiplexed RNA in situ Hybridization, bioRxiv (under review by Nature Communications).

CHAPTER 3: DART-FISH REVEALS CELLULAR HETEROGENEITY OF HUMAN CEREBRAL CORTEX

Introduction

Human brains are mainly composed of brainstem, cerebellum, thalamus, and cerebrum, with each structure governing different physiological functions. In cell biology, it has been reported that human brains contain approximately 100 billions neurons and are estimated to have 10 times more glia cells[1]. Strikingly, the cerebral cortices and cerebellum together make up to 92% of the mass and 99% of neurons in human brains[1, 142, 143]. Human cerebral cortex, especially, underwent the greatest expansion in size and elevated expression of specific genes during evolution, which distinguish us from other non-human primates[1, 4, 5]. To better study the complexed cerebral cortex, German anatomist Korbinian Brodmann pioneered the work and defined around 40 Brodmann areas based on the cytoarchitectural organization of neurons in each cortical areas[6]. Regardless of the original definitions of cellular organization, some Brodmann areas are closely related to various brain functions such as sensory, motor, and information processing functions. For example, Brodmann area 4 is linked to primary motor cortex (M1C) while area 17 is related to primary visual cortex (V1C). Nevertheless, the spatially resolved molecular underpinnings of these cortical regions are still elusive. The high correlation of the cytoarchitectures and these functionally defined areas drives the attention of many scientists to decipher the underlying cellular and transcriptional variations of the cortical sheet, which ultimately gives rise to complexed brain functions.

Early efforts to investigate the molecular phenotypes of cortex in a large-scale manner include the AIBS's Allen Brain Atlas that integrates extensive gene expression data, connectivity data and MRI images to dissect the molecular underpinnings

of mouse and human brain functions[55, 144]. However, the cellular organization and the molecular signatures were still obscured by the microarray measurement of cortical subdivisions. Although they tried to fill the gap by profiling approximately 1000 genes in human visual and temporal cortices based on ISH platform that measures one single gene per tissue section at a time, the work was still limited by the coverage and the multiplexity of in situ gene measurement [29]. Recently, highly multiplexed spatial transcriptomic methods have been widely developed and opened a new avenue to profile spatial distribution of diverse cell types and their molecular underpinnings in human cerebral cortex[87]. For example, Markus M. Hilscher and Mats Nilsson’s group used their HybISS method to map 120 genes and 75 transcriptomic cell types in human MTG[145]. Nevertheless, compared to the giant human cerebral cortical sheet, this work is still far from comprehensive. More recently, Jorstad et al. used snRNA-seq to profile the transcriptomic cytoarchitecture of 8 neocortical areas along the rostrocaudal axis, including ACC, PFC, M1C, S1C, MTG, A1C, AnG, and V1C, in human neocortex[126]. In this work, they profiled the whole transcriptomes of each single-cell with deep sampling of each neocortical area (80-100k nuclei per area), which allow the characterization of rare cell types such as L5 ET and SST CHODL neurons. Besides, they used MERFISH to validate spatial distribution of 140 genes in MTG and V1C. In their study, they revealed these cortical areas are consistently composed of 24 coarse cell subclasses. Among these areas, V1C showed the most drastic changes of the cellular organization, with substantial expansion of excitatory neurons in layer 4. Although this large-scale study provide refined cellular and molecular characterization in these widely sampled human neocortical areas, their discoveries are based on dissociated nuclei and the in situ gene expression measurement was only validated in MTG and V1C. Therefore, a spatially resolved cell atlas of more human cortical areas is still much needed.

In this chapter, we directly applied our DART-FISH method to comprehensively characterize the transcriptomic cell types and their distributions in 9 cerebral cortical areas. Here, we widely sampled the same 8 areas of human neocortex in

Jorstad et al.’s study and 1 FI area belonging to human allocortex. Since the sensitivity of DART-FISH is relatively moderate and the multiplexity is not transcriptome-wide, we spatially resolve 484 genes, including marker genes of those 24 coarse cell subclasses defined in Jorstad et al.’s study, and use these genes as anchors to align SNARE-seq2 nuclei in space by a computational tool, Tangram, developed in the BICCN consortium[146]. With Tangram alignment, the whole transcriptomes of SNARE-seq2 nuclei can now be analyzed in space. Together, the results in this chapter provide the quantitative cytoarchitectures and the spatial transcriptomics of 9 specialized areas in human cerebral cortex by integrative analysis of DART-FISH cells and SNARE-seq2 nuclei.

Transcriptomic profiling of single-cells in human cortical areas

Experimental workflow

To potentially capture the major transcriptomic variations of human cortical sheet, we sampled 8 neocortical areas along the rostrocaudal axis, including primary (M1C, S1C, A1C, and V1C) and association areas (DFC, ACC, MTG, and AnG) as well as 1 allocortical area, FI, that show drastic cytoarchitectural variations[147] (Figure 3.1a). The cortical areas were delineated in the postmortem human brains by anatomical landmarks and histological verification. Each area has 3-6 postmortem human brains for tissue dissection. The same tissue blocks were used for SNARE-seq2 and DART-FISH experiments. SNARE-seq2 assays started with isolated nuclei while DART-FISH assays were performed on cryo-sectioned tissue slices (Figure 3.1b). SNARE-seq2 assays were used to profile the whole transcriptomes of isolated nuclei derived from all cortical areas following the procedures described in Plongthongkum et al.’s paper[94]. DART-FISH assays were carried out on tissue sections having intact laminar layers as shown in Figure 3.2. SNARE-seq2 is able to profile the whole transcriptome of each isolated nuclei but lose the spatial information. In contrast,

DART-FISH can resolve the spatial localization of each gene but has moderate sensitivity and limited gene throughput. Therefore, a deep-learning-based computational tool, Tangram[146], was used to generate spatially resolved single-cell transcriptomic cell atlases across 9 cortical areas by aligning the SNARE-seq2 nuclei with DART-FISH cells via the similar expression of the common genes.

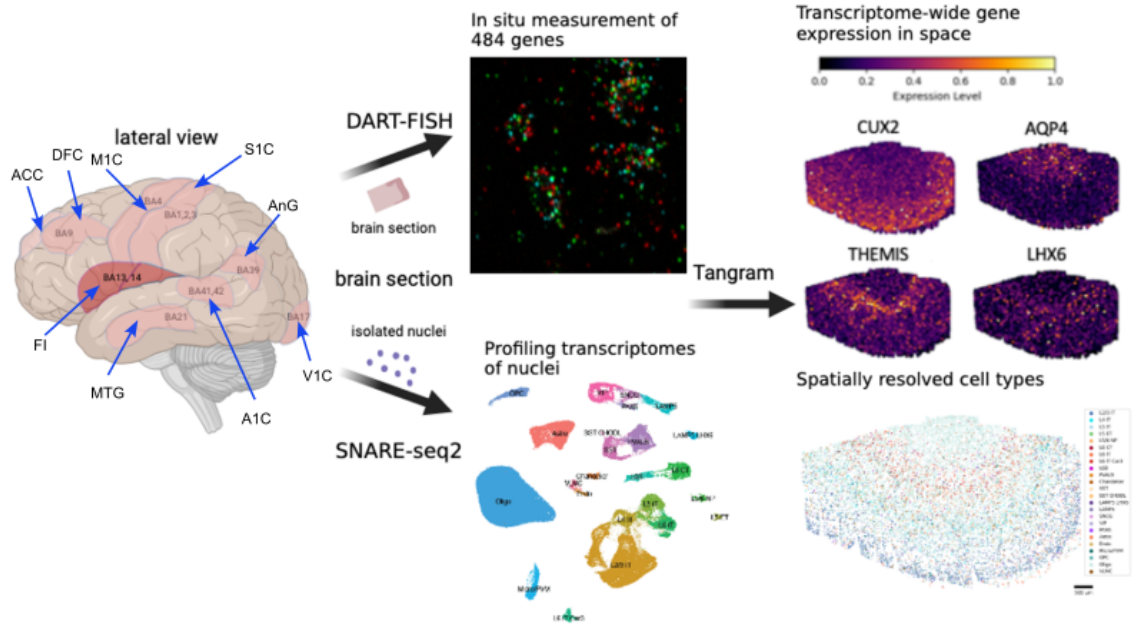


Figure 3.1 **DART-FISH and SNARE-seq assays on 9 cortical areas.** Dissected tissue blocks of 9 cortical areas that include ACC (BA 24), DFC (BA 9), FI (BA 13, 14), M1C (BA 4), S1C (BA 1,2,3), MTG (BA 21), A1C (BA 41, 42), AnG (BA 39), and V1C (BA 17), were collected. Then, DART-FISH assays measured 484 genes in situ while SNARE-seq2 profiled the whole transcriptomes of dissociated nuclei. Lastly, DART-FISH cells were aligned with SNARE-seq nuclei to obtain transcriptome-wide gene expression and cell types in space.

Transcriptome-wide gene expression profiling by SNARE-seq2

The total numbers of deeply sampled SNARE-seq2 nuclei across each cortical area is summarized in Table 3.1 and the numbers of nuclei derived from each human brain are summarized in Supplementary Figure S6a. In total, SNARE-seq2 assayed 1,733,910 nuclei that include 598,558 excitatory neurons, 246,337 inhibitory neurons, and 889,015 non-neuronal cells from multiple human brains. Approximately 140k-440k nuclei were sampled from each cortical area, which was considered deep

sampling (80k-100k) based on the standard mentioned in Jorstad’s paper[126]. These nuclei were annotated based on label transfer from snRNA-seq datasets published by AIBS[126]. The annotated nuclei were organized in 3 levels of hierarchies, including 3 classes (GABAergic, Glutamatergic, and Non-neuronal), 24 subclasses, and 150 cell types as illustrated in Supplementary Figure S7. Joint clustering of excitatory neurons, inhibitory neurons, and non-neuronal cells from each area was depicted in Supplementary Figure S8a, b, and c. At the subclass level, SNARE-seq2 nuclei were annotated with 9 excitatory neuron subclasses (L2/3 IT, L4 IT, L5, IT, L5 ET, L5/6 NP, L6 IT, L6 IT Car3, L6 CT, and L6B), 9 inhibitory neuron subclasses (PVALB, Chandelier, SST, SST CHODL, LAMP5 LHX6, LAMP5, SNCG, VIP, PAX6), and 6 non-neuronal subclasses (Astro, Endo, Micro/PVM, Oligo, OPC, VLMC). Besides, these 24 cell subclasses are consistently existing in all cortical areas, indicating these 24 cell subclasses are highly consistent cellular components of these 9 cortical areas (Supplementary Figure S8d, e, and f). Also, the subclass level annotation reflects some biological meanings. For example, the IT, CT, and NP excitatory neurons reflect different patterns of axon projections and the L2/3, L4, L5, and L6 reflect their laminar layer positions. Besides, the PVALB, SST, VIP, LAMP5 are the neuropeptides secreted by different inhibitory neurons. Hence, we decided to select marker genes to uncover the transcriptomic cell types at the subclass level for the DART-FISH experiments.

Table 3.1 **Deep sampling of areal nuclei by SNARE-seq2**. Each area contains nuclei isolated from multiple donors. Nuclei belong to the same area are pooled together to achieve deep sampling of each cortical area. The total numbers of nuclei represent those SNARE-seq nuclei passing RNA QC.

Area	Donors	Total number of nuclei
ACC	H19.30.001, H19.30.002, H19.30.004	146,798
DFC	H19.30.001, H19.30.002, H19.30.004	235,659
FI	H19.30.001, H19.30.002, H19.30.004, UW7118	345,506
M1C	H19.30.001, H19.30.002, H19.30.004, H18.30.001, H18.30.002	445,918
S1C	H19.30.001, H19.30.002, H19.30.004, UW7118	281,456
MTG	H19.30.001, H19.30.002, H19.30.004, UW7118	223,637
A1C	H19.30.001, H19.30.002, H19.30.004, UW7118	229,276
AnG	H19.30.001, H19.30.002, H19.30.004	386,299
V1C	H19.30.001, H19.30.002, H19.30.004, UW7118	284,256

Measurement of 484 genes in situ by DART-FISH

DART-FISH assays were carried out on the tissue sections derived from parts of the same tissue blocks as SNARE-seq2 datasets. To subsequently annotate 24 cell subclasses in each DART-FISH, we used gpsFISH to analyze SNARE-seq2 datasets and select 484 informative genes for cell annotation[148]. Experimentally, DART-FISH assays were performed on 2 parallel tissue sections from a brain and 2 brains were used to represent each cortical area (2 replicates per tissue block and 2 tissue blocks per area). To achieve wide sampling of these cortical areas, we imaged centimeter scale ROI ($\sim 23 - 65mm^2$) to measure 15k-47k cells in each DART-FISH dataset.

However, the integrity of tissue sections from each tissue blocks varied. So, an extra parallel section from the same tissue block was collected for Nissl staining to delineate the laminar layers of each tissue section (Figure 3.2). Only DART-FISH datasets from tissue section having L2-L6 laminar layers were used for Tangram alignment and downstream analysis. It's worth mentioning that these DART-FISH datasets were generated by an early version of DART-FISH having lower sensitivity (see detailed steps in attached "Chen_BICCN-DART-FISH-protocol.pdf"). Hence, a subset of marker genes in the 484-gene panel were used for *ad hoc* cell annotation. The marker genes were selected based on the combination of differential analysis of SNARE-seq2 datasets and literature curated markers. Specifically, differentially expressed genes having > 2 fold-change, > 2 -fold expression in a specific cell subclass than other cell subclasses, and included in the DART-FISH 484-gene panel were considered marker genes for cell annotation. With these criteria, we obtained 163 marker genes across 24 cell subclasses in total and the lists of marker genes for each cell subclasses were summarized in Table 3.2.

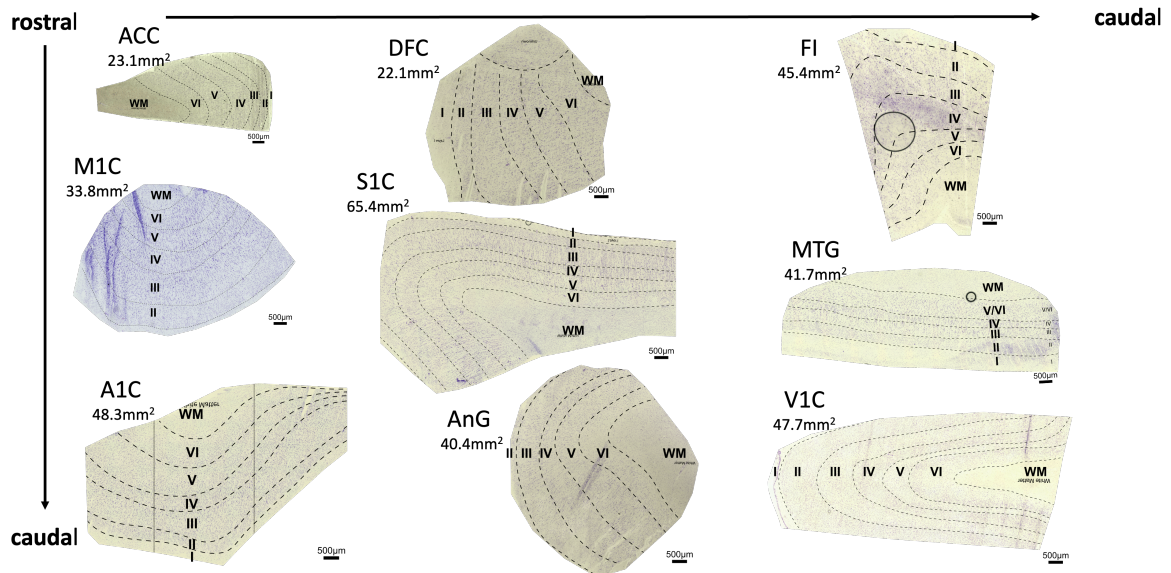


Figure 3.2 Nissl staining of 9 human cortical areas along rostrocaudal axis. The areas from rostral to caudal axis include ACC, DFC, FI, M1C, S1C, MTG, A1C, AnG, and V1C. Numbers under area indicates the size of ROI of each area. WM: White matter

Table 3.2 Marker genes to annotate 24 cell subclasses of DART-FISH cells.

Class	Subclass	marker genes
Excitatory	L2/3 IT	<i>CUX2, HS6ST3, LINGO2, CBLN2, DGKB, DLGAP2, EPHA6, RGS6, KCNQ5, LINC00507</i>
Excitatory	L4 IT	<i>CNTN5, DCC, IL1RAPL2, PLCH1, POU6F2, RORB, TRHDE, CPNE4, TSHZ2</i>
Excitatory	L5 IT	<i>IL1RAPL2, CLSTN2, CPNE4, CADPS2</i>
Excitatory	L5 ET	<i>SORCS1, COL24A1, VAT1L, TMTC1, CBLN2, ZNF804B</i>
Excitatory	L5/6 NP	<i>HS3ST4, KCNB2, TRHDE, TSHZ2, ZNF385D, DCC, ZNF385B, ADAMTSL1, SORCS2, CPNE4, NXPH2, TOX</i>
Excitatory	L6 IT	<i>CLSTN2, MCTP1, ZNF804B, SLC35F3, RALYL</i>
Excitatory	L6 IT Car3	<i>CPNE4, GALNT14, NWD2, SEMA6D, SORBS2, ZNF804B, CADPS2, CLSTN2, SATB2, TMEM132D, TRHDE, FRAS1</i>
Excitatory	L6 CT	<i>ADAMTSL1, CLSTN2, COL24A1, HS3ST4, KIAA1217, MCTP1, SEMA3E, RALYL, SULF1, ZNF385B</i>
Excitatory	L6B	<i>ADAMTSL1, HS3ST4, KIAA1217, MCTP1, ROBO2</i>
Inhibitory	PVALB	<i>BTBD11, GAD2, GRIP1, KCNC2, KIAA1217, MYO16, NXPH1, PTPRM, SLIT2, ZNF385D, PAM, SGCZ, TENM3, UNC5D, EPHA6, SAMD5, ERBB4, SOX6, PVALB</i>
Inhibitory	Chandelier	<i>ANK1, FSTL5, GABRG3, GRIP1, MYO16, PLCXD3, THSD7A, TMTC1, ZNF385D, CNTN5, FRMPD4, KCNC2, PTCHD4, TMEM132D, ZNF536, PVALB, COL15A1</i>

Table 3.2 **Marker genes to annotate 24 cell subclasses of DART-FISH cells (Continued).**

Class	Subclass	marker genes
Inhibitory	SST	<i>GRIK1, GRIN3A, GRIP1, KIAA1217, NXPH1, SYNPR, TENM3, TRHDE, PLCH1, XKR4, ZNF385D, PTPRM, SOX6, COL19A1, ROBO2, PCDH15, SST</i>
Inhibitory	SST CHODL	<i>SST, NXPH2, NDST4, NOS1, TRPC6, LRP8, GRIK1, SOX6, TMTC1, IL1RAPL2, RUNX1T1, PAM, TOX, GRIP1, PTPRT, UNC5D, COL24A1, TENM3, KCNB2, XKR4</i>
Inhibitory	LAMP5 LHX6	<i>ADARB2, FSTL5, GAD2, GRIN3A, GRIP1, KCNC2, MGAT4C, MYO16, NTNG1, NXPH1, PDGFD, PTCHD4, PTPRT, SEMA6D, SGCZ, TOX, MTUS2, UNC5D, SLIT2, ZNF385D, KIRREL3, SLC6A1, LAMP5, LHX6</i>
Inhibitory	LAMP5	<i>ADARB2, CXCL14, FSTL5, GAD2, GRIK1, GRIN3A, GRIP1, KCNC2, MYO16, NXPH1, PTCHD4, PTPRM, PTPRT, RAB3C, RELN, SGCZ, SLC6A1, ZNF385D, CCDC85A, FRAS1, MTUS2, ZMAT4, CNTN5, MGAT4C, NXPH2, TMEM132D, ADAMTSL3, LAMP5</i>
Inhibitory	SNCG	<i>ADARB2, CNR1, CXCL14, FSTL5, GRIK1, RELN, ZNF385D, CNTN5, IL1RAPL2, SORCS1, GRIP1, THSD7A, ADRA1A</i>
Inhibitory	VIP	<i>ADARB2, CNR1, SYNPR, THSD7A, ROBO2, ERBB4, ROBO1, VIP</i>

Table 3.2 **Marker genes to annotate 24 cell subclasses of DART-FISH cells (Continued).**

Class	Subclass	marker genes
Inhibitory	PAX6	<i>ADARB2, CCDC85A, CNR1, CXCL14, GRIP1, MYO16, NXPH1, RELN, ZNF385D, IL1RAPL2, GABRG3, CNTN5, PLD5, CDH4, FSTL5, ROBO2, ANGPT1</i>
Non-neuronal	Astro	<i>ADGRV1, ATP1A2, ETNPPL, GFAP, GPC5, KCNN3, RYR3, SLC1A2, IQCA1, SPARCL1, NEAT1, FGFR3, FBXL7, AQP4</i>
Non-neuronal	Oligo	<i>BCAS1, CD22, MOBP, MOG, NCKAP5, PDE1C, PIEZO2, PLP1, RNF220, ZNF536, OPALIN, TMTC2, ALCAM, QKI</i>
Non-neuronal	OPC	<i>CA10, COL11A1, DCC, LHFPL3, MAML2, NXPH1, PCDH15, PDGFRA, PTPRZ1, SEMA5A, SOX6, VCAN, XYLT1, NTNG1, TNS3, KCND2</i>
Non-neuronal	Micro/PVM	<i>APBB1IP, HS3ST4, LRMDA, LRRK1, LYN, MAF, MAML2, MRC1, SFMBT2, ST6GAL1, TNS3, IL6ST, ZFHX3, ARHGAP12</i>
Non-neuronal	Endo	<i>BTNL9, FLT1, ST6GAL1, LRMDA</i>
Non-neuronal	VLMC	<i>ATP1A2, COL1A2, DCN, SLC6A1, COL4A2, PLCE1, ZFHX3</i>

Alignment of DART-FISH cells with pre-annotated SNARE-seq nuclei

To align DART-FISH cells with pre-annotated SNARE-seq2 nuclei including 24 cell subclasses, we ran Tangram alignment in the "cells" mode[146]. This provided us the probabilistic mapping of SNARE-seq2 nuclei in space based on the cosine simi-

larity of the expression of those 484 genes in DART-FISH cells. To do a deterministic cell annotation, we normalized the Tangram probabilities of each cell subclass and transferred the cell subclass label, having the highest normalized Tangram probability, to the DART-FISH cells. Nonetheless, this cell-by-cell mapping between individual DART-FISH cells and SNARE-seq2 nuclei was noisy. We applied additional filtering by the count of the marker genes in each annotated DART-FISH cells. The marker genes for each subclass used in this filtering step were summarized in Table 3.2. Pre-annotated DART-FISH cells with low count of subclass marker genes were excluded for downstream analysis and the numbers of total annotated cells in each area were summarized in Table 3.3. Besides, each individual DART-FISH cell was further aligned with the SNARE-seq2 nucleus, having the highest Tangram probability among that specific cell subclass. Ultimately, the single-cell transcriptomes were spatially resolved and spatially resolved cell atlases of these cortical areas were generated.

Table 3.3 Annotation of DART-FISH cells in each cortical area. DART-FISH cells were aligned with pre-annotated SNARE-seq2 nuclei with 24 subclasses. Tangram[146] was used to align the DART-FISH cells with SNARE-seq2 nuclei having the highest cosine similarity of the gene expression profile. DART-FISH cells with low count of transcripts, low Tangram probabilities, and low marker gene count were filtered. The numbers of annotated cells after filtering in each DART-FISH dataset are summarized here.

Area	Dataset ID	ROI (mm^2)	Total number of DART-FISH cells	Total number of annotated cells	Annotation rate
ACC	20221005	23.05	15,385	9,378	61.0%
DFC	20221201	22.08	15,453	7,881	50.4%
FI	20221017	45.39	33,448	18,961	56.7%
M1C	20220912	33.84	30,797	15,690	50.9%
S1C	20220909	65.44	47,821	26,125	54.6%
MTG	20220725	41.65	36,683	14,701	40.1%
A1C	20220928	48.33	35,852	20,084	56.0%
AnG	20221107	40.45	35,501	17,961	50.6%
V1C	20220712	47.71	33,448	27,026	80.8%

Spatially resolved cell atlas by DART-FISH

Spatial distribution of cell subclasses in human cortical areas

The spatial distributions of the 24 cell subclasses in each cortical area were illustrated in Figure 3.3, 3.4, and 3.5. In Figure 3.3, the excitatory neurons showed pronounced spatial gradient along the cortical depth, with L2/3 IT neurons in the upper layers and L6 CT as well as L6B neurons in the deeper layers. It's worth noting that most cortical areas revealed a noticeable and similar band of L4 IT neurons in the middle layer, except the rostral most ACC and the caudal most V1C. Based on histological staining, L4 IT-like neurons are lost in agranular ACC and M1C. By our molecular cell-typing method, we truly observed L4 IT neurons in the middle layer of ACC even though they are more widely distributed and can spread to deeper L2/3 and upper L5 layers. Also, we observed an obvious band of L4 IT neurons in the middle layer of M1C even though other studies reported the lack of L4 IT neurons having L4 neuron-like morphological characteristics in agranular M1C.[113, 149–151]. Thus, our molecular cell-typing method can indeed uncover previously missing cell types defined by histological staining. In addition, V1C showed drastic expansion of L4 IT neurons that spread into L2/3 IT neurons dominating layers. Also, in the middle of L2/3 and L4, the L2/3 IT neurons and L4 IT neurons were well mixed together. This observation is in line with Jorstad et al.'s study that relied on Nissl staining to dissect laminar layers in V1C[126]. Consistent with our results, they also observed L4 IT neurons in specialized layers 4A and 4B. Moreover, they also found multiple types of L2/3 IT neurons in layers 4A and superficial part of layer 4B, which may explain the intermingled pattern of L2/3 IT and L4 IT neurons in our results.

In contrast to the discernible gradient distribution of excitatory neurons across the cortical depth, inhibitory neurons generally have wider distribution in all cortical areas, as shown in Figure 3.4. This may be due to their developmental origins outside the cortical plate[128, 129]. Still, we observed those MGE-derived inhibitory neurons (PVALB, Chandelier, SST, and SST CHODL) were mostly localized in the deeper

layers while the CGE-derived inhibitory neurons (LAMP5 LHX6, LAMP5, SNCG, VIP, and PAX6) were generally localized in the upper layers. The PAX6 inhibitory neurons, especially, were positioned at more superficial layer than upper layer 2/3 in most cortical areas. This may due to the transcription factor *PAX6* highly expressed in their apical progenitor (AP) cells in VZ not only controls the neurogenesis of PAX6 inhibitory neurons but also mediates their migration to the superficial cortical layers during corticogenesis[152–154].

Lastly, the non-neuronal cell subclasses were mostly sporadically distributed across cortical depth, except Oligo and Astro cells (Figure 3.5). We found that the Oligo cells were highly enriched in the WM zone and spread into the cortical layer 4 even though their progenitor cells, OPCs, were uniformly distributed in all cortical areas. [134]. This may reflect their important roles in myelination and axon support in the WM zone[134, 155]. Astro cells, in contrast, show uniform with slight depletion in the middle layers in most cortical areas such as M1C, S1C, A1C, and V1C. The wide distribution of Astro cells may be related to their supporting roles in regulatory functions such as metabolic, homeostatic, neurotransmitter clearing, and blood-brain barrier permeability in the human brain[156, 157]. To sum up, Oligo cells demonstrate more similar but pronounced laminar positioning in all cortical areas while Astro cells are widely distributed but have more differences in laminar organization across cortical areas.

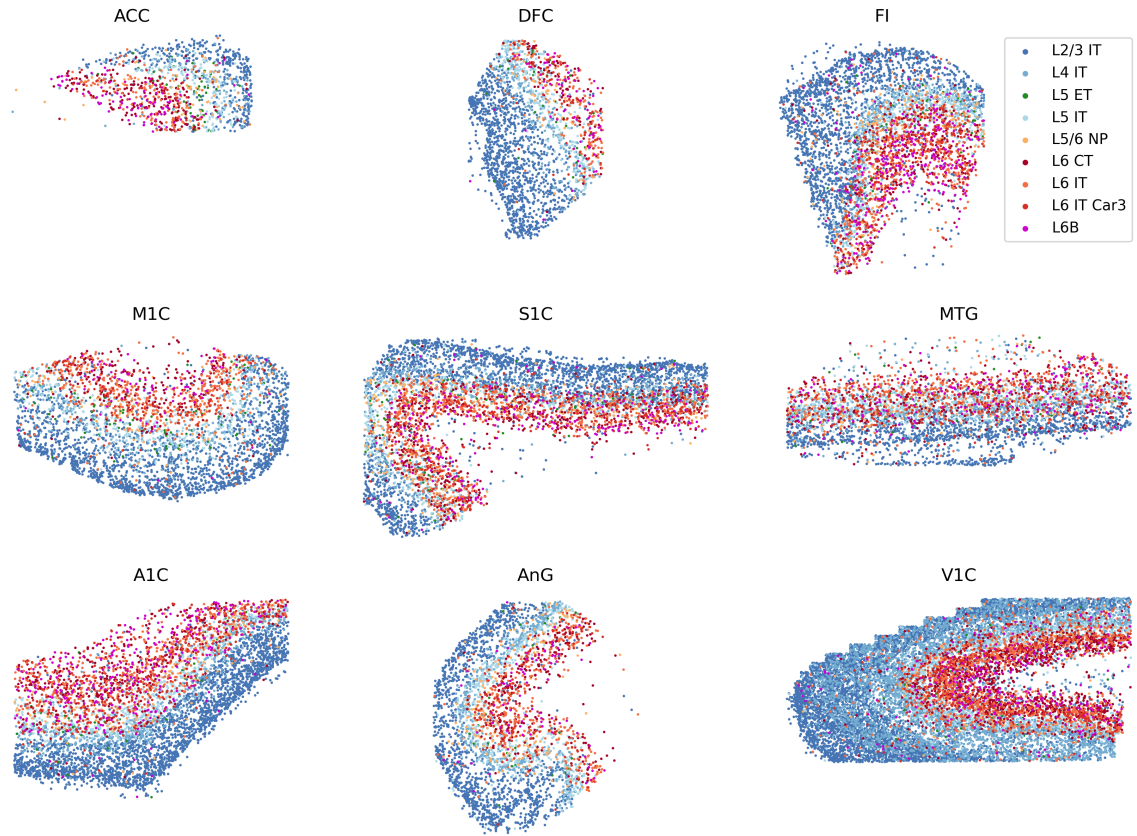


Figure 3.3 **Cell atlas of excitatory neurons in 9 areas of human cerebral cortex.** Spatial distribution of excitatory neurons, including L2 IT, L4 IT, L5 IT, L5 ET, L5/6 NP, L6 IT, L6 IT Car3, L6 CT, and L6B, across 9 cortical areas were depicted.

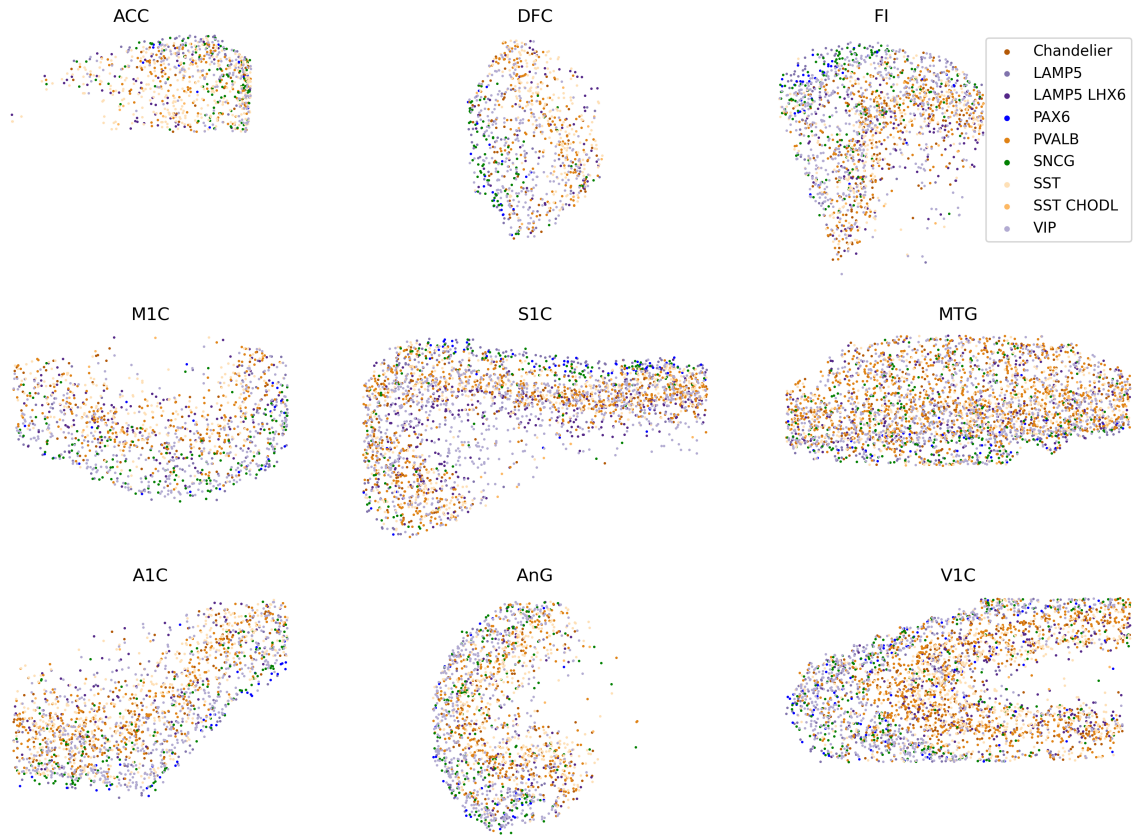


Figure 3.4 Cell atlas of inhibitory neurons in 9 areas of human cerebral cortex. Spatial distribution of inhibitory neurons, including PVALB, Chandelier, SST, SST CHODL, LAMP5 LHX6, LAMP5, SNCG, VIP, and PAX6, across 9 cortical areas were depicted.

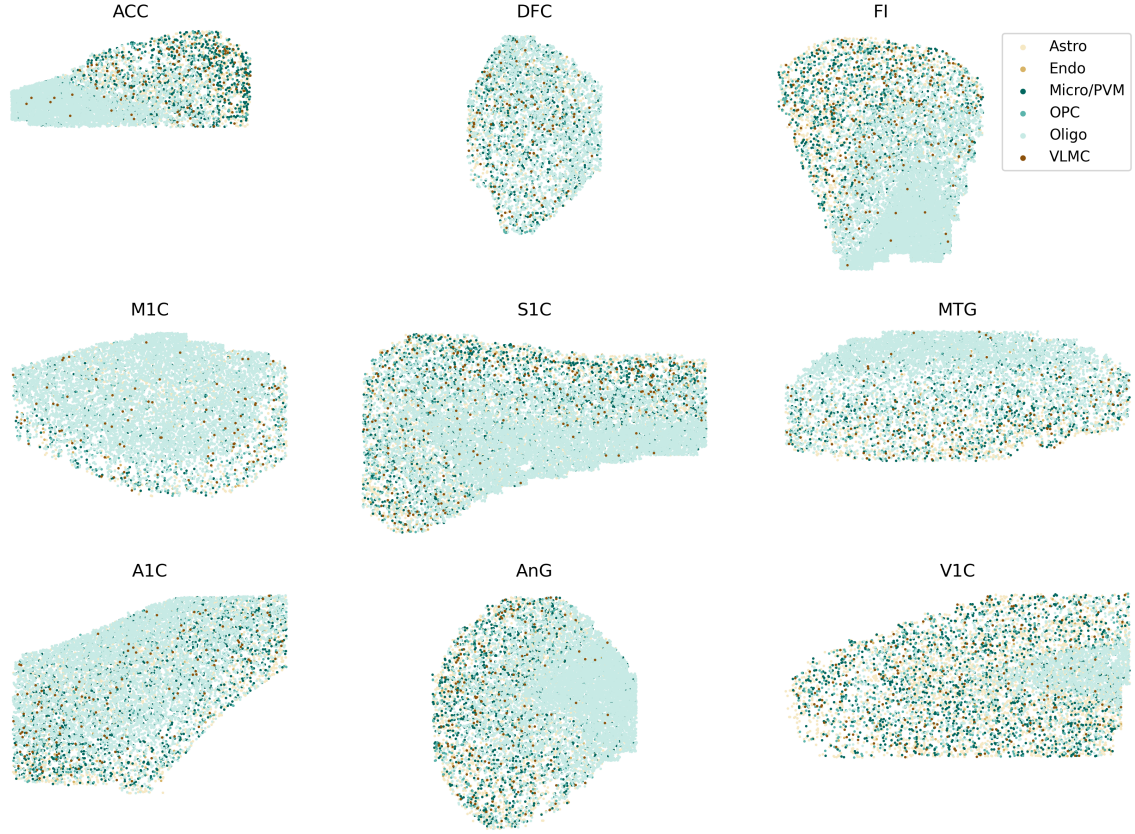


Figure 3.5 **Cell atlas of non-neuronal cells in 9 areas of human cerebral cortex.** Spatial distribution of non-neuronal cells, including Astro, Endo, OPC, Oligo, Micro/PVM, and VLMC across 9 cortical areas were depicted.

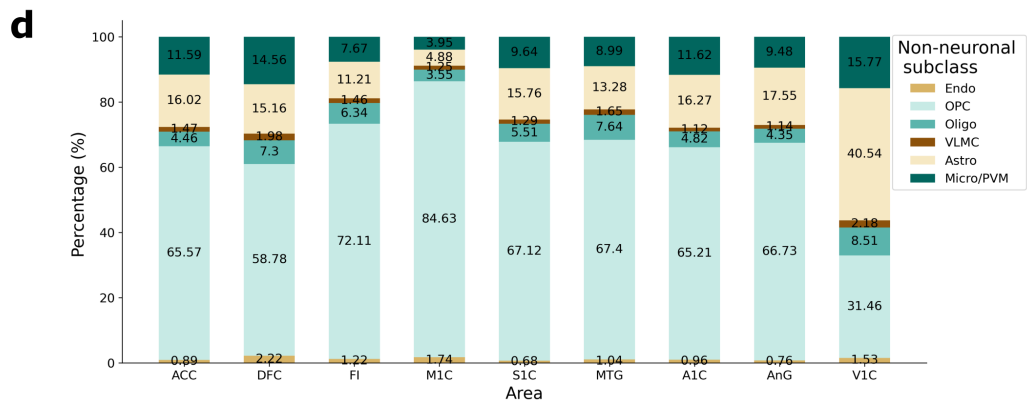
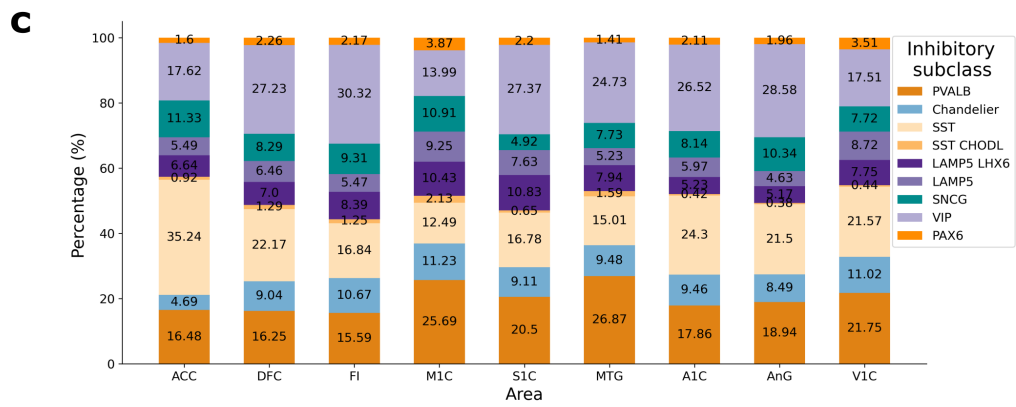
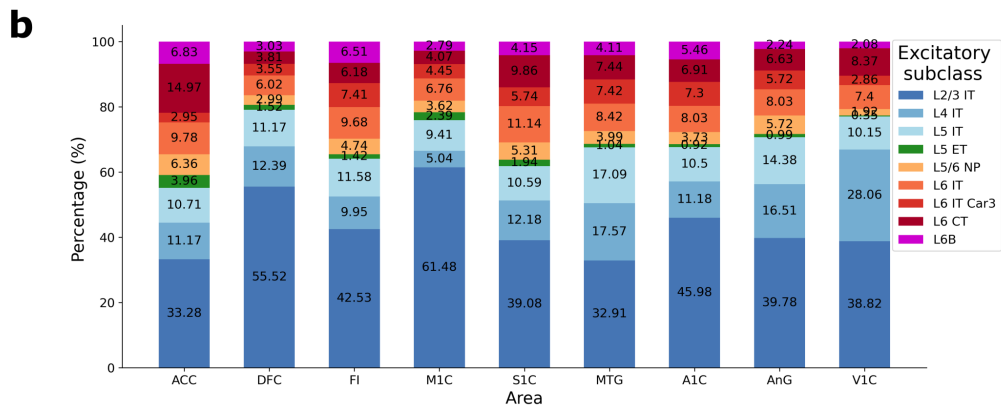
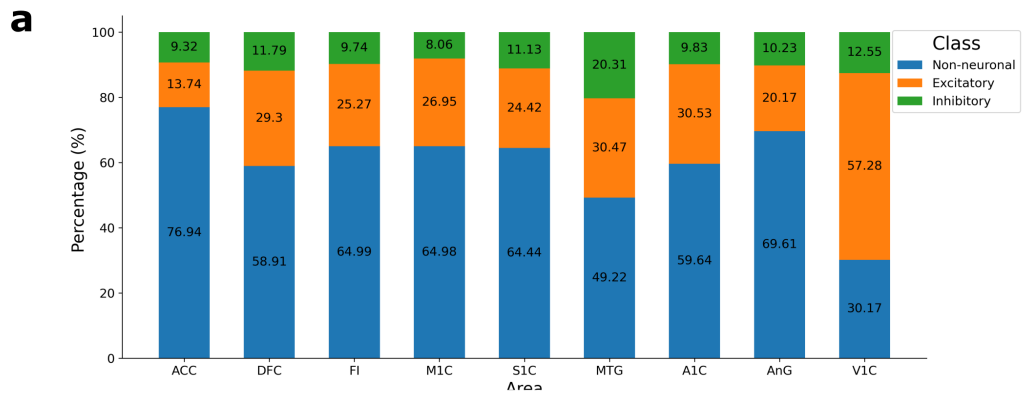
Cell composition changes reveal arealization of human cerebral cortex

The compositions of 24 cell types across 9 cortical regions were summarized in Figure ???. The numbers and proportion of non-neuronal cell may vary across tissue sections because some tissue sections had smaller WM zone (DFC, V1C) during tissue dissection as proved by Nissl staining (3.2). Also, each cortical area may not have enough sampling based on the deep sampling standard (80k-100k per area) mentioned in Jorstad et al.'s paper[126], which may cause the missing of some rare cell types such as SST CHODL neurons. Hence, we mainly compared our discoveries of the excitatory and inhibitory neurons at class and subclass level in all cortical areas but the numbers of non-neuronal cells were still summarized in Figure ??d.

At the cell class level, we noticed most cortical areas, except ACC and V1C,

contain 70-80% excitatory neurons and 20-30% inhibitory neurons, respectively, consistent with the numbers of cortical neurons discovered in previous studies[23]. In rostral most ACC, the excitatory neurons slightly decrease to 60% while 82% cortical neurons are excitatory in the caudal most V1C, showing a rostral-low-caudal-high cellular composition of the excitatory neurons in Figure ??a. At the cell subclass level, excitatory neuron subclasses still exhibit the most areal differences. The majority of excitatory neurons belongs to L2/3 IT neurons in all areas. Consistent with their spatial distribution (Figure ??b) and other studies, L4 IT neurons have lower abundance in ACC and M1C[113] while substantial increase of L4 IT neurons was observed in V1C, in line with the drastic expansion of layer 4 in V1C. On the other hand, L5 ET, L5/6 NP and L6B neurons showed the highest abundance in ACC and the portion of these neurons generally decreased along the rostrocaudal axis. Strikingly, even without deep sampling, we still observed the highest abundance of a relatively rare cell types L5 ET in ACC and gradually decreased along the rostrocaudal axis. L5 ET neurons have subcortical projecting axons, different electrophysiological features and gene expression patterns[113, 158]. The least abundance of L5 ET neurons in V1C is in line with previous studies in macaque, discovering the axon projections to subcortical targets were derived from sparse neurons (i.e. Meynert cells) positioned in deep layers in V1C[159, 160]. For inhibitory neurons, the compositions of inhibitory neuron subclasses are generally similar across all cortical areas (Figure ??c). Also, the inhibitory neurons of MGE origin roughly constitute 50-60% of cortical interneurons while CGE-derived inhibitory neurons constitute 30-40%, consistent with previous report[34]. Together, these findings indicate the cellular organization exhibit strong areal distinction at the most rostral and caudal axis while the cytoarchitectures of cortical areas in the middle of the rostrocaudal axis are surprisingly similar.

Figure 3.6 Proportion of cell types at class and subclass level across 9 cortical areas. (a) Percentage of excitatory, inhibitory, and non-neuronal cell class in ACC, DFC, FI, M1C, S1C, MTG, A1C, AnG, and V1C. (b) Percentage of excitatory neuron subclasses across 9 cortical areas. (c) Percentage of inhibitory neuron subclasses across 9 cortical areas. (d) Percentage of non-neuronal subclasses across 9 cortical areas.



Along with the areal changes of excitatory neuron subclass, we also found the areal differences of excitatory and inhibitory neuron ratio (E:I ratio) across cortical areas. As depicted in Figure 3.7, the E:I ratio of most cortical areas fluctuated around 2-3, which is lower than widely reported 5:1 E:I ratio in mouse. Also, E:I ratio in the rostral most ACC is slightly lower but the E:I ratio in the caudal most V1C tremendously increase to 4.5:1, which is highly likely due to the expanded layer 4 in V1C. This finding, again, highlights the extreme specializations of V1C in human cerebral cortex.

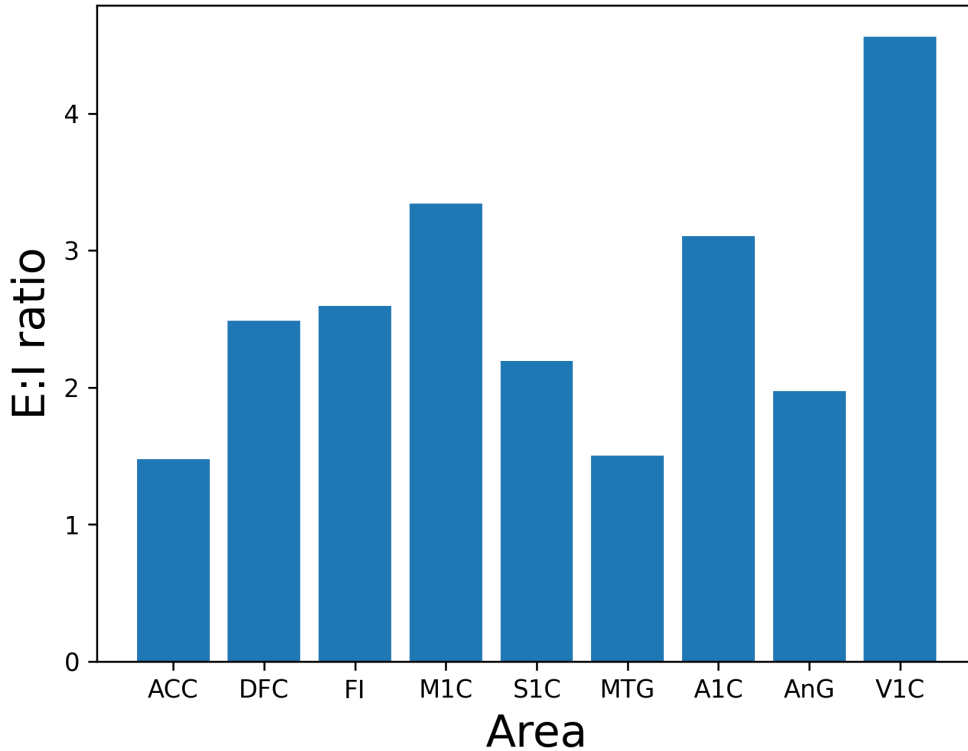


Figure 3.7 **E:I ratio across 9 cortical areas along the rostrocaudal axis.** E:I ratio represents the ratio of excitatory neurons and inhibitory neurons.

Transcriptomic landscape of human cerebral cortex

To examine the similarity of transcriptomes in aligned DART-FISH cells across multiple cortical areas, we embedded these neurons and non-neuronal cells in com-

mon latent space and visualize their transcriptomic similarity in UMAPs. Since these aligned DART-FISH cells derived from multiple human subjects, we corrected the batch effect caused by human subjects prior to UMAP embedding with Harmony[161](Right panel of Figure 3.8). Then, the cells in UMAPs are colored by subclass and cortical area. We found that excitatory neurons seem to show subtle transcriptomic gradient along the rostrocaudal axis and the caudal most area, V1C, has more specialized L2/3 IT, L4 IT, and L6 CT neurons that have rather different gene expressions from the counterparts from other cortical areas (Figure 3.8a, b). Also, we notice that the excitatory neurons in the middle areas, M1C and S1C, are mostly intermixed, indicating that these two physically proximal areas have similar transcriptomic landscapes even they are related to distinct functions.

In contrast, the MGE-derived and CGE-derived inhibitory neurons generally have very similar transcriptomes, as demonstrated by the intermixed cells across cortical areas in Figure 3.8c, d. Similarly, transcriptomes in non-neuronal cells show no pronounced areal difference, as exhibited by the intermingled cells in Figure 3.8. Therefore, these observations point out that the areal difference of transcriptomic landscapes is mostly driven by excitatory neurons even though the transcriptomic difference only shows pronounced differences at the rostral most and caudal most areas even in excitatory neurons. The results are consistent with the previous large-scale studies based on microarray and bulk-RNA-seq[53–56]. Combined our findings with previous studies, it further indicates that transcriptomic landscape alone is insufficient to define the functionally discrete areas.

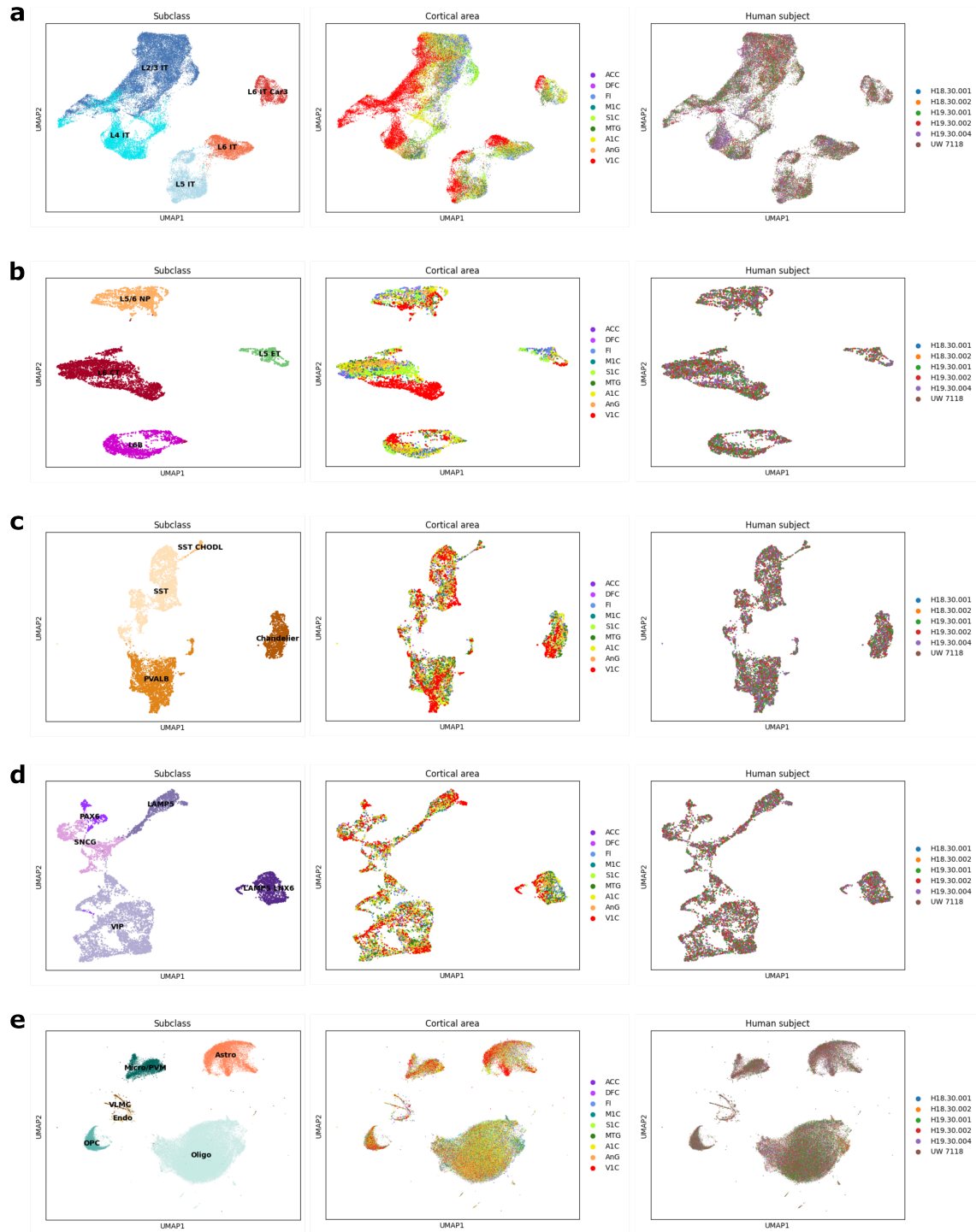


Figure 3.8 UMAPs of excitatory neurons, inhibitory neurons, and non-neuronal cells across 9 cortical areas. (a) UMAP embedding of IT-projecting excitatory neurons. (b) UMAP embedding of non-IT projecting excitatory neurons. (c) UMAP embedding of MGE-derived inhibitory neurons. (d) UMAP embedding of CGE-derived inhibitory neurons. (e) UMAP embedding of non-neuronal cells.

Discussion

The adult human cerebral cortex is extremely complexed and can be parceled into many functionally discrete areas. Deciphering the cellular organization and the molecular phenotypes under these functional areas was heavily investigated in neuroscience. To see if these discrete areas also have areal specializations at single-cell and molecular level, BICCN started extensive characterizing the M1C in human, mouse, as well as marmoset[113] and uncover over 100 transcriptomic cell types organized in three levels of hierarchies, which lay a good foundation to cross-areal studies of human cerebral cortex. Jorstad et al. from AIBS used snRNA-seq to profile the whole transcriptomes of isolated nuclei from 7 more human cortical areas and leveraged the 24 subclasses discovered in BICCN’s work to annotated other areas, which ultimately ended up with a comprehensive cross-areal consensus taxonomy[126]. Although this large-scale study shed light on the transcriptomic cytoarchitectures of human cerebral cortex, snRNA-seq doesn’t preserve the spatial information and they only validated the spatial distribution of 140 genes and cell subclasses in V1C and MTG by MER-FISH. Thus, spatially resolved transcriptomic cell atlases of these cortical areas are still missing.

To fill the gap, we provided the spatially resolved cell atlases and the underlying molecular underpinnings of 9 cortical areas (8 neocortical areas and 1 allocortical area) in human brain in this chapter. By aligning DART-FISH cells and pre-annotated SNARE-seq2 nuclei with Tangram, spatially resolved transcriptomes at single-cell level across these cortical areas were generated. We found these previously defined 24 cell subclasses in M1C also resided in all other cortical areas, implying generalized cellular components of human cerebral cortex. If we quantify these cellular components at class and subclass level, we noticed that the proportion of excitatory neurons generally increase along the rostrocaudal axis. Among these cortical neurons, excitatory neurons generally show noticeable spatial gradients along the cortical depth and have pronounced changes of cell composition along the rostrocaudal axis. For example, L4

IT neurons generally increased along the rostrocaudal axis except the agranular area ACC and M1C while L5 ET exhibited an inverse pattern and had the least abundance in the caudal most V1C. In contrast, inhibitory neurons generally had similar cellular abundance across all cortical areas and their spatial distribution show wider gradients, except PAX6 interneurons which are constantly localized in the superficial layer. Specifically, MGE-derived interneurons (PVALB, Chandelier, SST, SST CHODL) were positioned in deeper layers while CGE-derived interneurons (LAMP5 LHX6, LAMP5, SNCG, VIP, and PAX6) were localized in upper layers. Besides, the E:I ratio is critical to maintain proper functions of neuronal circuits and the imbalance of excitatory and inhibitory neurons is commonly related to neuropsychiatric diseases[162]. Here, we quantified the E:I ratio in all cortical areas of healthy human brain and noticed that the E:I ratio in most area fluctuated between 2 and 3 with the rostral most ACC having slightly lower E:I ratio and the caudal most V1C having drastically higher E:I ratio at 4.56. This is highly likely due to the substantial expansion of layer 4 and increase of L4 IT neurons in V1C. Additionally, non-neuronal cells were mostly sporadically distributed across the cortical areas except Oligo cells. Oligo cells were highly enriched in the WM zone, gradually decreased along the cortical layers, and were depleted in the upper cortical layers. This spatial pattern of Oligo cells was consistent in all cortical areas, reflecting their important roles in myelination and axon support in the WM zone. On the other hand, we found the transcriptomic landscapes of inhibitory neurons and non-neuronal cells demonstrated high similarity across these cortical areas. Only excitatory neurons exhibited subtle transcriptomic gradients along the rostrocaudal axis and the caudal most area, V1C, has specialized L2/3 IT, L4 IT, and L6 CT neurons having more transcriptomic differences than their counterparts in other areas.

Altogether, these findings highlight the similar cellular organization among the middle cortical areas along the rostrocaudal axis. Only the rostral most ACC and the caudal most V1C show pronounced specializations, especially V1C having the largest extent of difference in cell composition, cytoarchitecture, and E:I ratio to

other cortical areas. These results are consistent with other large-scale studies by performing microarray, bulk RNA-seq and sc/snRNA-seq on microdissected adult human brain tissue[53–56, 126]. Also, several time course studies of human cortex had discovered the dynamic transcriptomes showed the most pronounced areal distinction in early brain development but the differences of the spatial transcriptomes decrease in the adulthood. This implies the discrete functional differences across these cortical areas in adult human brain cannot be sufficiently explained solely by their spatial transcriptomes. Future study integrating spatial transcriptomes with functional properties such as electrophysiological features and connectivity is needed to decipher if the human cortical areas lie on a molecular continuum or discrete zones to mirror those functionally discrete areas.

Acknowledgement for Chapter 3

I want to acknowledge people involved in this large-scale cell atlasing work. The work in Chapter 3 is part of a big collaborative work between the Zhang lab, the Chun lab, and the Ren lab. The pre-annotated SNARE-seq2 datasets were totally generated by Dr. Jinghui Song. Also, all human brain sections and isolated nuclei were prepared by Dr. Carter Palmer from the Chun lab. I generated all DART-FISH datasets from 9 human cortical areas. Then, I aligned DART-FISH datasets with nuclei from SNARE-seq2 datasets using the 484 genes in DART-FISH datasets as anchors. Subsequently, quantification and downstream analysis were performed and exhibited in this chapter. Deeper analysis will be done, and this collaborative work will be published in the future. I am the primary author of this chapter.

Chapter 3 is based on the following work which will be published in the future:

Chen, Chien-Ju; Song, Jinghui; Palmer, Carter; Zemke, Nathan; Yang, Bing; Diep, Dinh; Lake, Blue B.; Zhang, Yida; Kharchenko, Peter V.; Ren, Bing; Chun, Jerold; and Zhang, Kun. BICCN work under preparation for publication.

CHAPTER 4: Conclusions and perspectives

In the first part of this dissertation (Chapter 2), we developed DART-FISH, a highly multiplexed RNA in situ hybridization method, to measure gene expression in the spatial context of human tissues, having high autofluorescence background[89–93]. The early crosslinking of cDNA to PA gel immediately after reverse transcription showed 1.5-fold DART-FISH signals and the built-in RiboSoma labeling more effectively delineated the boundaries of cell body than traditional nuclei staining. Also, the autofluorescence in human tissues was overcome by simple RCA reaction that amplifies DART-FISH signals by over 100-fold. Besides, the fast decoding combined with the computational framework, SpD, allowed scalable measurement of large human tissues with 20x objective in a shorter timeframe. As a proof of principle, we showed its reproducibility and benchmarked its sensitivity and specificity in human M1C and compared it to RNAscope and other methods such as MERFISH and EELFISH. Besides, DART-FISH measurement of 121 genes in human M1C successfully recapitulated the cytoarchitectures of 20 of 24 previously identified cell subclasses[113]. This proving its direct applicability to postmortem human brain tissue sections, which lay a great foundation to generate human cerebral cortex atlas in Chapter 3. In addition to its application to map the cytoarchitecture of human M1C as described in Chapter 2, a few other future directions will be mentioned here to incorporate more modalities into this streamlined DART-FISH workflow.

First, we can apply DART-FISH to measure more cells in thick tissue blocks so that deep sampling of that tissue block can potentially collect more numbers of rare cell types for statistical analysis. This requires DART-FISH to combine the tissue clearing procedures and light-sheet microscopy to reduce scattering light and imaging time during image acquisition[163, 164]. Second, if subcellular localization of RNAs in finer structures such as dendritic spines and branches is required in a research study, we can combine DART-FISH with expansion microscopy to spatially resolved gene expression and these finer structures[141]. Third, if we want to include membrane

protein receptors or extracellular matrix proteins to investigate cell-cell interactions in healthy and pathological niches, antibody staining can be easily incorporated into the DART-FISH workflow[140].

In the second part of this dissertation (Chapter 3), we demonstrated the application of DART-FISH to uncover the cellular heterogeneity of 9 human cortical areas, ACC, DFC, FI, M1C, S1C, MTG, A1C, AnG, and V1C, along the rostrocaudal axis. Combined with the transcriptome profiling of single-nuclei in these areas by SNARE-seq2, we generated spatially resolved transcriptomic cell atlases of these human cortical areas. This large-scale study is trying to address a long-lasting biological question - if the functionally discrete areas in human cerebrum are mirrored by continuum or discrete zones of cortical sheet at cellular and molecular levels. In our findings, we realized the adult cortical sheet were all made up of 24 cell subclasses belonging to excitatory neurons, inhibitory neurons, and non-neuronal cells. Also, the cytoarchitectures of these areas were mostly similar, except the two poles at the rostral and caudal axis. The rostral most ACC was slightly specialized by fewer excitatory neurons, wider distribution of L4 IT neurons, and lower E:I ratio at 1.47. In contrast, the most caudal V1C showed drastic specialization, manifested by the higher abundance of excitatory neurons especially L4 IT neurons, pronounced expansion of layer 4, and higher E:I ratio at 4.56 than other middle areas (E:I ratio 2-3). Similarly, in the transcriptomic landscape V1C also have specialized L2/3 IT, L4 IT, and L6 CT neurons that show more transcriptomic differences. However, the transcriptomic landscapes in most cell subclasses are generally similar, except excitatory neurons which also show subtle transcriptomic gradients across the rostrocaudal axis. These findings are consistent with other large-scale studies based on microarray, bulk RNA-seq, and sn/scRNA-seq[53–56, 126]. Thus, it seems that the areal transcriptomic distinction is subtle unless in the polar areas, ACC and V1C.

Although the results of this large-scale study only seem to recapitulate the subtle areal difference of the transcriptomic landscape in the cortical sheet, bridging the results with other omic data such as the chromatin landscape in the cortical sheet

may help to explain the functional features of human cerebral cortex[45]. In a recent study of *Drosophila*, people discovered some spatial transcription factors influenced the chromatin landscape, which opened specific chromatin domains for temporal transcription factors to bind, and ultimately generated diverse neuronal cell types[165]. We may need to incorporate chromatin accessibility data in future work to decipher the roles of gene regulatory networks in producing areal molecular signatures and diverse cell types across cortical areas. Another particularly interesting and possibly more relevant aspect is to incorporate spatial transcriptome with neuronal connectivity, since similar cellular organization can still give rise to distinct functional areas by wiring the neurons in different ways. Nonetheless, integration of spatial transcriptome and neuronal connectivity may require further tool development to map these two modalities together in a large-scale manner.

To sum up, the development of DART-FISH allows us to uncover the heterogeneity of complexed human cerebral cortex at multiple scales spanning from molecular, cellular to the tissue level. The workflow is so simple that allow easy incorporation of other modalities. I am excited about tailoring this method to address many biological questions in future perspectives.

MATERIALS AND METHODS

Reagents and enzymes

Reagents and enzymes were listed as name (supplier, catalog number) in the Supplementary file, Chen_Supplementary Table 1. Also, the human brain probes used in chapter 2 were listed in Chen_Supplementary Table 2.

Post-mortem human brain tissue

De-identified postmortem human brain tissue was collected by AIBS after obtaining the consents from the decedent's legal next-of-kin. Brain tissue collection was performed in accordance with the provisions of the United States Uniform Anatomical Gift Act of 2006 described in the California Health and Safety Code section 7150 (effective 1/1/2008) and other applicable state and federal laws and regulations. The Western Institutional Review Board (WIRB) reviewed the use of de-identified post-mortem brain tissue for research purposes. Male and female donors of 18-68 year-old with no known history of neuropsychiatric or neurological conditions were included for the postmortem human brain tissue collection. Meanwhile, the blood samples of the donors were screened for infectious disease (HIV, Hepatitis B, and Hepatitis C) and donors negative for all three infectious diseases were included. A few brain tissue samples were obtained from the University of Washington Biorepository and Integrated Neuropathology (BRaIN) Laboratory under UW School of Medicine and HIPAA compliance. Informed consent was obtained for the use of data and samples.

Three donor brains with post-mortem interval ≤ 12 hours and RIN score ≥ 7 were selected for DART-FISH assay (Table S1). Regions were identified and isolated utilizing architectural landmarks, aided by the Allen Brain Human Brain Atlas [147]. Multiple parallel 10- μ m-thick cryosections were taken from the tissue block and mounted onto vectabond (Vector Laboratories, SP-1800-7)-coated 24 x 60 mm No.1.5 coverslips (Azer Scientific, 1152460). Brain cryosections were stored at -80 °C until use.

Table S1 **Metadata of human brain donors.** Human brain tissue blocks were collected by Allen Institute of Brain Science and University of Washington.

Subject ID	Gender	Age (years)	Post-mortem interval (PMI)	Hemisphere
UW7118	Female	27	8.2 h	N/A
H19.30.001	Male	42	8 h	right
H19.30.002	Male	29	7.5 h	right

Padlock probe production

Gene selection

In Chapter 2, a list of genes was selected based on differential expression analysis of snRNA-seq data from human M1C data[64, 113] and a few curated marker genes were added manually to target 121 genes in human M1C. In Chapter 3, a 484-gene panel was selected mostly by gpsFISH[148] while a few curated marker genes were added manually to uncover 24 subclasses across excitatory neurons, inhibitory neurons, and non-neuronal cells.

DART-FISH probe design

For short genes (length < 1.5kb), we defined the constitutive exon as the union of all isoforms in GencodeV41. For other genes, the constitutive exon was defined as regions in RefSeq where at least 33% of isoforms share. We used a modified version of ppDesigner[110] to find padlock target sequences along the constitutive exons. ppDesigner was run on two settings: 1) no overlap between probes allowed, 2) overlap of up to 20nt allowed. Individual target sequences were constrained between 17nt and 22nt long with the total target sequences ≤ 40 nt. The resulting target sequences were aligned to GRCh38/hg38 with BWA-MEM[166] and sequences with MAPQ < 40 or secondary alignment were removed. We further removed probes that have GATC

(DpnII recognition site). A maximum of 50 probes per gene were selected, prioritizing the non-overlapping set. Finally, the target sequences were concatenated with amplification primer sequences, universal sequence, and gene-specific decoder sequences to produce final padlock probe sequences (Supplementary Figure S1) and were ordered as an oligo pool from Twist Bioscience. the brain probe set was amplified with AP1V7U and AP2V7 primer pair (Chen_Supplementary Table 1).

To select a set of barcodes, we computationally created all possible barcodes in the compact format: a n digit barcode with “1”, “2” and “3” representing each of the three fluorescent channels and “0” indicating ”off” cycles. For example, the barcode for *RORB* in Figure 2.1c is “132000” in the 4-digit format. This amounted to 480 multi-color barcodes for brain. We then used a brute force algorithm to find the largest subset of barcodes, $\mathbf{Q}$, in which every pair had a Hamming distance ≥ 2 . Followed by this, we created a graph, $\mathbf{G}$, in which every possible barcode is a node, and pairs of nodes are connected with edges if their Hamming distance is 1. We then found a maximal independent set (MIS, networkx v2.6.2) that included the nodes in $\mathbf{Q}$. This method ensured that every pair of barcodes in the MIS have Hamming distance > 1 . Because the algorithm for finding MIS is random, we ran it 20,000 times and selected the largest MIS across the runs. For the brain, the MIS consisted of 159 barcodes, 121 of which were randomly assigned to the genes.

TBU gel

All TBU gels in padlock probe production were made with the home-made recipe from the lab. Briefly, a master mix of 6% TBE-Urea denaturing gel casting solution was prepared based on the recipe in Table 1. Then, we filtered the solution with sterile 30 um Celltrics filter (Sysmex, 04-004-2326) and add it into individual empty gel cassettes, 1.0 mm (ThermoFisher Scientific, NC2010). 1-well or 12-well combs were inserted to the top of the gel cassettes. Place the gel cassettes in the fume hood until the gel formed. The TBU gels were stored at a 4°C fridge until use.

Table S2 **TBE-Urea denaturing gel (TBU gel) casting solution recipe**. All volumes are in mL.

reagent	15 mL	24 mL	30 mL	48 mL
8M Urea solution	11.25	18	22.5	36
10x TBE buffer	1.5	2.4	3	4.8
40% Acrylamide/Bis solution (29:1)	2.25	3.6	4.5	7.2
10% Ammonium persulfate	0.1	0.15	0.2	0.3
TEMED	0.005	0.0075	0.01	0.015
Total	15.15	24.225	30.3	48.45

Mass padlock probe production

A step-by-step protocol can be found in the Supplementary file, [Chen_Enzymatic-padlock-probe-preparation.pdf](#) and is illustrated in Supplementary Figure S1. Briefly, oligo pools were PCR amplified on a 96-well plate (10pM per reaction) using KAPA SYBR fast and 0.4 μ M of each amplification primer (AP1V7U and AP2V7 for brain, Chen_Supplementary Table 1, Supplementary Figure S1) until plateau. The PCR products were pooled and concentrated with ethanol precipitation and further purified using QIAquick PCR purification kit (Qiagen, 28106).

The purified amplicons were divided into parallel reactions (about 5 μ g each) and were digested with Lambda Exonuclease (0.5U/ μ L) in 1x buffer (NEB, M0262L) at 37°C for 2 h and purified using Zymo ssDNA/RNA clean & concentrator kit following manufacturer's instructions (Zymo, D7011). Next, the single-stranded probes were further digested with 5 units of USER enzyme (NEB, M5505L) in 1x DpnII buffer at 37 °C for 3 h. Subsequently, for each reaction we added DpnII guide oligo (Supplementary Figure S1) to final concentration of 5uM in 1x DpnII buffer, heated the mix to 94°C for 2 min, cooled to 37°C and added 50 units of DpnII in 1x DpnII buffer and incubated for 5 h. Finally, probes were size-selected using the TBE-Urea gel.

DART-FISH sample preparation

The general workflow, including reverse transcription, cDNA crosslinking, padlock probe capture, RCA, colony crosslinking and image acquisition, is illustrated in Figure 2.1. A step-by-step protocol can be found in Supplementary files (Chen.DART-FISH-paper-protocol.pdf, Chen.BICCN-DART-FISH-protocol.pdf).

Notes: The DART-FISH protocol evolved during my graduate study. In Chapter 2, I crosslinked cDNA molecules and embedded them in gel right after reverse transcription (Chen.DART-FISH-paper-protocol.pdf) while I crosslinked colonies and embedded colonies in gel after RCA reaction in Chapter 3 (Chen.BICCN-DART-FISH-protocol.pdf).

Reverse transcription and cDNA crosslinking

Tissue sections were fixed in 4% PFA in 1x PBS at 4 °C for 1 h, followed by two 3-minute washes with PBST (1x PBS and 0.1% Tween-20). Then, a series of 50%, 70%, 100%, and 100% ethanol were used to dehydrate the tissue section at room temperature for 5 min each. Next, tissues were air-dried for 5 minutes and in the meantime, silicone isolators (Grace Bio-Labs, 664304) were attached around the tissue sections. Then, the tissue sections were permeabilized with 0.25% Triton X-100 in PBSR (1x PBS, 0.05U/ μ L Suprase In, 0.2U/ μ L Enzymatics RNase Inhibitor) at room temperature for 10 min, followed by two pre-chilled PBSTR (1x PBS, 0.1% Tween-20, 0.05U/ μ L Suprase In, 0.2U/ μ L Enzymatics RNase Inhibitor) washes and a water wash. Next, the sections were digested with 0.01% pepsin in 0.1 N HCl (pre-warmed 37 °C for 5 min) at 37 °C for 90 seconds and washed with pre-chilled PBSTR twice. Afterwards, acrydite-modified dT and N9 primers (Acr_dc7-AF488_dT20 and Acr_dc10-Cy5_N9) were mixed to a final concentration of 2.5 μ M with the reverse-transcription mix (10U/ μ L SuperScript IV (SSIV) reverse transcriptase, 1x SSIV buffer, 250 μ M dNTP, 40 μ M aminoallyl-dUTP, 5 mM DTT, 0.05U/ μ L Suprase In and 1U/ μ L Enzymatics RNase inhibitor), incubated at 4 °C for 10 min and then transferred to a humidified 37 °C oven for overnight incubation. After reverse tran-

scription, tissue sections were washed with pre-chilled PBSTR twice and incubated in 0.2 mg/mL Acryloyl-X, SE in 1x PBS at room temperature for 30 min. Then, the tissue sections were washed once with PBSTR, followed by incubation with 4% acrylamide solution (4% acrylamide/bis 37:1, 0.05U/ μ L Suprase-In, and 0.2U/ μ L RNase inhibitor) at room temperature for 30 min. Subsequently, the acrylamide solution was aspirated and gel polymerization solution (0.16% APS and 0.2% TEMED in the 4% acrylamide solution) was added. Immediately, the tissues were covered with Gel Slick (Lonza #50640)-treated circular coverslips of 18 mm diameter (Ted Pella, 260369), transferred to an argon-filled chamber at room temperature and incubated for 30 min. After gel formation, the tissue sections were washed with 1x PBS twice and the coverslip was gently removed with a needle. At this point, the cDNA is crosslinked to the polyacrylamide gel.

Padlock probe capture

After cDNA crosslinking in gel, remaining RNA was digested with RNase mix (0.25U/ μ L RNase H, 2.5% Invitrogen RNase cocktail mix, 1x RNase H buffer) at 37 °C for 1 h, followed by two PBST washes, 3 min each. Then, the padlock probe library was mixed with Ampligase buffer. Then, the mixture was heated to 85 °C for 3 min and cooled on ice. Then, the mixture was supplemented with 33.3U/ μ L such that the final concentration of padlock probe library was 100 nM in 1x Ampligase buffer. Subsequently, the samples were incubated at 37 °C for 30 min, and then moved to a 55 °C humidified oven for overnight incubation.

RCA and rolon crosslinking

After padlock probe capture, the tissue sections were washed with 1x PBS three times, 3 min each and hybridized with RCA primer solution (0.5 μ M `rca_primer`, 2x SSC, and 30% formamide) at 37 °C for 1 h. Then, the tissue sections were washed with 2x SSC twice and incubated with Phi29 polymerase solution (0.2 U/ μ L Phi29 polymerase, 1x Phi29 polymerase buffer, 0.02 mM aminoallyl-dUTP, 1 mg/mL BSA, and 0.25 mM dNTP) at 30 °C in a humidified chamber for 7 h. Afterwards, the tissue

sections were washed with 1x PBS twice, 3 min each and the colonies were crosslinked with 5 mM BS(PEG)9 in 1x PBS at room temperature for 1 h. The crosslinking reaction was terminated with 1M Tris, pH 8.0 solution at room temperature for 30 min. Finally, samples were washed with 1x PBS twice and stored in a 4 °C fridge until image acquisition.

Image acquisition

Human brain tissue samples were stained with 1x TrueBlack in 70% ethanol at room temperature for 2 min to lessen the lipofuscin autofluorescence and washed with 1x PBS three times for 3 min each before imaging. For anchor round imaging, a mixture of anchor round probes, including DARTFISH_anchor_Cy3, dcProbe10_ATT0647N, and dcProbe7_AF488 probes, were diluted to 500 nM in 2x SSC and 30% formamide. Then, the samples were stained with anchor round probes at room temperature for 5 min and washed with 1 mL washing buffer (2x SSC, 10% formamide and 0.1% Tween-20) twice for 2 min each prior to imaging. The samples were immersed in 1 mL imaging buffer (2x SSC and 10% formamide) during imaging. For decoding imaging, each imaging cycle started with incubating samples with stripping buffer (2x SSC, 80% formamide, and 0.1% Tween-20) at room temperature for 5 min, washed with washing buffer twice for 2 min each, stained with a mixture of AlexaFluor488, Cy3, and ATTO647 fluorophore-labeled decoding probes (dcProbe0-AF488, dcProbe0-Cy3, and dcProbe0-ATTO647N as an example for round 1) for 10 min, and washed with washing buffer three times for 2 min each. Then, the samples were immersed in 1 mL of imaging buffer while imaging. After the last cycle of decoding imaging, DRAQ5 staining (5 μ M, room temperature, 10 min) was performed for nuclei segmentation. Z-stack images were acquired by Leica TCS SP8 confocal microscope with 20x oil-immersed objective (NA 0.75) and pixel size of 284 nm x 284 nm and 1024 x 1024 pixels per image.

DART-FISH data processing workflow

The DART-FISH datasets were processed by our home-built pipeline. Briefly, raw z-stack images with 4 channels (3 fluorescent channels and brightfield) from the microscope were registered to a reference round by affine transformation implemented in SimpleElastix[167] using the brightfield channel as the anchor. Then, each field of view (FOV) underwent decoding to obtain a list of candidate spots. Spots from all FOVs were pooled and filtered (See Sparse deconvolution decoder for more details). To obtain the global position of the colonies, the FOVs were stitched by applying FIJI's[168] Grid/Collection Stitching plugin[169] (in headless mode) to the registered and maximum-projected brightfield images. Note that the theoretical positions of the FOVs, defined by the microscope, were used as initial positions for stitching. Cell boundaries were segmented with Cellpose (v2.1.1)[170, 171]. Note that Cellpose's "cyto" was fine tuned on each tissue by manually segmenting a handful of composite images DRAQ5 (nuclei channel) and N9 cDNA stain or RiboSoma (cyto channel) using the package's graphical user interface.

SparseDeconvolutionDecoder

In DART-FISH, each gene is represented by a barcode that can be read out in n rounds of 3-channel imaging. Each barcode is designed to emit fluorescence (be "on") in exactly k rounds, each time in a single fluorescent channel and stay "off" in other rounds. We concatenate the rounds and channels and represent the barcodes as $3n$ -dimensional vectors. In other words, barcode i is represented by vector x_i in which 1's are placed where "on" signal is expected, and 0's everywhere else. The codebook matrix X ($3n \times N$) is then defined as $X = [x_1, x_2, \dots, x_N]$, where N is the total number of barcodes. In the same way, for every pixel we concatenate the fluorescent intensity values (scaled between 0 and 1) to create a $3n$ -dimensional vector.

The fluorescence signal at each pixel can be sourced from more than one colony if the distance between neighboring colonies is smaller than the optical resolution of the imaging system, or if 3-dimensional stacks are analyzed as maximum-projected

2D images. Nevertheless, because of physical constraints, only a handful of colonies are expected to be the source of signal to each pixel. In this regard, because of the redundancy in the barcode space, combinations of barcodes in one pixel can be decomposed into their original composing barcodes. We formulated this problem as a regularized linear regression problem where a weighted sum of a few barcodes creates the observed signal intensity, where the vector shows the contribution of each barcode (Figure 2.1d) with most w_s ($1 \leq s \leq N$) elements equal to 0. We initially used lasso to solve this problem ($\alpha' = 0$ in Figure 2.1d) to promote the sparsity of w , but later decided to use elastic net with a non-zero value for α' that is much smaller than α ($\alpha'/\alpha = 0.01$) to increase stability. We call the solution to this problem $\hat{w}_{lasso}$. Note that, we constrain the problem to positive weight values ($\hat{w}_{lasso_s}$ for every s). The regression problems are solved for all the foreground pixels ($\|y\| > 0.25$) individually. For every barcode i , we can construct an image with the estimated weight values as pixels: 0 for background and rejected pixels, and non-zero values from $\hat{w}$. We call these images weight maps. Figure 2.1f shows weight maps constructed with $\hat{w}_{lasso}$ which have not been filtered. With our current barcode space, we can only confidently decompose bi-combinations, so for every instance of the elastic net problem, we applied an elbow filter and accepted the solution only when the top one or two weights were significantly larger than other weights. In more details, for every pixel, the weights in $\hat{w}_{lasso}$ are sorted in decreasing order. If the second largest weight is smaller than half of the top weight, then the top weight passes the elbow filter. Otherwise, if the third largest weight is smaller than 30% of the largest weight, the top two weights pass the elbow filter. All the values that do not pass the filter are set to zero. For accepted solutions, we performed an ordinary least square (OLS) regression using the top one or two weights to obtain unbiased weights ($\hat{w}_{OLS}$). Supplementary Figure S3d shows weight maps constructed with $\hat{w}_{OLS}$ (OLS maps) after applying a Gaussian smoothing.

Estimating channel-specific coefficients

So far, we have assumed that pixel intensities from different rounds and fluorescent channels all have the same scale and distribution. However, there is usually a variation among rounds and fluorescent channels, with some channel-rounds being brighter than others. To account for this effect, we model the channel-specific variations as a multiplicative factor that connects the weights at each pixel to intensities: $y = c \odot Xw$ where $c = [c_1, c_2, \dots, c_{3n}]^T$ is the channel coefficient vector and $\odot$ denotes element-wise multiplication. Suppose for a set of pixels $y^{(1)}, y^{(2)}, \dots, y^{(P)}$ the true barcode weights $w^{(1)}, w^{(2)}, \dots, w^{(P)}$ are given. For pixel i and channel j , we could write: $y_j^{(i)} = c_j \sum_{b=1}^N X_{jb} w_b^{(i)} = c_j \sum_{b=1}^N (x_j)_b w_b^{(i)}$ where $(x_j)_b$ shows the b 's element of the j 's barcode. In this case, each c_j can be estimated by solving an OLS problem between $y_j^{(\cdot)}$ and $\sum_{b=1}^N (x_j)_b w_b^{(\cdot)}$. Conversely, if the channel coefficients are given, we can set up the decoding problem with normalized intensities: $\bar{y} = y/c = Xw$ with $/$ being element-wise division. We estimate the channel coefficients in an iterative manner following the algorithm below:

- (1) Initialize $c = 1$ (no channel variation)
- (2) Take a random sample of foreground pixels
- (3) Normalize the pixel intensities in the sample with c
- (3) Run sparse deconvolution decoder on the normalized pixels
- (4) Keep pixels with one dominant unsaturated weight (weight in range 0.1 and 0.5) and obtain unbiased weights through OLS
- (5) Update the values of c by solving $3n$ OLS problems
- (6) Repeat step (3)- step (6) n_{iter} times

We do this procedure for 2 iterations and apply the obtained values when decoding all fields of view.

Setting the elastic net regularization parameter

Because of physical constraints, the solution to the deconvolution problem must be sparse, i.e., only a few non-zero weights should explain the observed intensities.

The sparsity of the solution is directly controlled by the L1 regularization term, α (Figure 2.1d). For a given pixel y , higher values of α shrink the estimated weights ($\|\hat{w}_{lasso}\|_1 \rightarrow 0$). Conversely, lower values of α allow more weights to be non-zero and $\|\hat{w}_{lasso}\|_1$ to grow larger. In fact, one can show if the L2 regularization term, α' , the largest weight to be undetected for a pixel made purely from one barcode is $w_{max} = \frac{3n}{k}\alpha$ [172]. For instance, given $\alpha = 0.05$ and codebook parameters $n = 6, k = 3$, then $w_{max} = 0.3$. This means that a pixel composed of one barcode needs to have an underlying intensity > 0.3 to get a non-zero $\hat{w}_{lasso}$. In other words, setting α too strictly will result in dimmer pixels to have $\hat{w}_{lasso} = 0$, while setting α too loosely will result in spurious non-zero values in $\hat{w}_{lasso}$ for brighter more complex pixels, potentially not passing the elbow filter and thus $\hat{w}_{OLS} = 0$. To accommodate a wide range of rolonity intensities, we choose adaptively based on the pixel norm $\|y\|_2$. First, we form a training data from a random subset of foreground pixels indexed by i . For a given pixel norm u , we find the alpha that maximizes a weighted sum of $\|\hat{w}_{OLS}^{(i)}\|_1$ giving more weights to training pixels with closer norms to u (equation $\star$):

$$\alpha(u) = \operatorname{argmax}_x \sum_i g\left(\frac{u - \|y^{(i)}\|_2}{\sigma}\right) \|\hat{w}_{OLS}^{(i)}(\alpha)\|_1$$

Where $g(\cdot)$ is the Gaussian function. In practice, for the training pixels we solve the sparse decoding problem for every value of α on a grid from 0.01 to 0.1 with a step size of 0.005, α_{train} , to obtain estimated weights $\hat{w}_{OLS}^{(\alpha)}$. Then we create a grid of norms u_{train} , spanning 0 and 2.8 with 50 steps. For every value of u in u_{train} , we solve equation $\star$ on the α_{train} grid. In other words, we create a lookup table connecting values of u_{train} to the best α in α_{train} . For new pixels, α is determined by the closest norm in the lookup table.

Spot calling

To call spots, Gaussian smoothing is applied to individual OLS maps, followed by `peak_local_max` filter (scikit-image 0.19.3[173]) which returns a binary image with 1's at the local maxima of the smoothed OLS maps. These peaks are then used as

markers for watershed segmentation. From each segmented region, these features are kept and used downstream: area, centroid, maximum and average intensity. This formed a list of candidate spots from each FOV.

Spot filtering

To control the specificity of the decoding procedure, we augmented the code-book with a number of barcodes (5-10% of the used barcodes) not used in the probe set (empty barcodes). After spot calling, we record the properties (e.g., area, maximum and average intensity) of spots with an empty barcode. Indeed, we see that empty spots tend to be smaller with lower average/maximum weight (Supplementary Figure S3f and g). On a small fraction of spots from all FOVs, we train a random forest classifier (scikit-learn v1.1.3) with area, maximum and average weights as features to predict empty/non-empty labels (Supplementary Figure S3h). We applied the classifier to all spots and obtained emptiness probabilities and set a threshold on these probabilities (0.3-0.35).

Spot assignment to cells

The cell boundaries were computed by applying `find_boundaries` (scikit-image 0.19.3[173]) to the segmentation mask. The distances of all spots were calculated to the closest cell boundary. The distance was set to 0 if a spot was inside a boundary. A spot was assigned to its closest cell if the distance was less than or equal to 3 μm for non-MBP and 0 μm for MBP spots in the brain.

RNAscope

Sample preparation

RNAscope HiPlex 50x probe stocks of human *SLC17A7*, *RELN*, *CUX2*, *RORB*, *CBLN2*, *FEZF2*, *GAD2*, *PVALB*, *LAMP5*, *PLP1*, *AQP4*, and *APBB1IP* with HiPlex12 Reagent Kit v2 (488, 550, 650) Assay (ACD, 324419) were purchased from Advanced Cell Diagnostics (ACD). The 50x probe stocks and RNAscope HiPlex diluent were warmed at 40 °C for 10 min. The pre-warmed 50x probe stocks were pooled and di-

luted to 1x with pre-warmed RNAscope HiPlex diluent before use. RNAscope experiments were carried out according to the manufacturer's protocol (document number UM324419) with slight modifications for the post-mortem human brain tissue. Briefly, the human brain tissue sections were fixed with 4% PFA in 1x PBS at 4 °C for 1 h and dehydrated with a series of 50%, 70%, 100%, and 100% ethanol at room temperature for 5 min each. Then, silicone isolators of 20 mm in diameter (Grace Bio-Labs, 664304) were applied around the tissue sections and the tissue sections were slightly digested with 5 drops of Protease IV at room temperature for 30 min and washed with 1x PBS for 2 min twice. Subsequently, enough volume of 1x pooled probes was added to cover the tissue sections entirely and the probe hybridization was performed in the 40 °C HybEZ Hybridization System for 2 h. Then, the tissue sections were washed with 1 mL 1x wash buffer at room temperature for 2 min twice. Later, the tissue sections were hybridized with RNAscope HiPlex Amp1, incubated in the 40 °C HybEZ Hybridization System for 30 min, and washed with 1x wash buffer at room temperature for 2 min twice. Afterwards, we followed the same process to hybridize the tissue sections with RNAscope HiPlex Amp2 and RNAscope HiPlex Amp3. Finally, we incubated the tissue section with freshly prepared 5% HiPlex FFPE reagent at room temperature for 30 min and washed the tissue sections with 1 mL 1x wash buffer at room temperature for 2 min twice prior to image acquisition.

Image acquisition

The tissue sections with silicone isolators were mounted above the objective of the confocal microscope and 4 cycles of imaging were performed to image 12 RNA species. In the first imaging cycle, RNAscope HiPlex Fluoro T1-T3 probes were pre-warmed at 40 °C, added to cover the tissue sections entirely, and hybridized with the tissue sections for 5 min thrice. After probe hybridization, the tissue sections were washed with 1 mL 1x wash buffer at room temperature for 2 min twice and immersed in 1 mL 4x SSC buffer. Z-stack images were acquired by Leica TCS SP8 confocal microscope with 63x oil-immersed objective (NA 1.4) and pixel size of 113 nm x 113 nm with each z-step at 1 μ m. Then, the fluorophores were cleaved with freshly

prepared 10% cleaving solution (100 μ L cleaving solution diluted with 900 μ L 4x SSC buffer) at room temperature for 15 min and the tissue sections were washed with 0.5% PBST (1x PBS with 0.5% Tween-20) at room temperature for 2 min twice. Repeat the fluorophore cleaving process once to ensure the fluorophores were removed entirely. Later, we repeated the same process to hybridize RNAscope HiPlex Fluoro T4-T6, RNAscope HiPlex Fluoro T7-T9, and RNAscope HiPlex Fluoro T10-T12 and image the corresponding RNA targets subsequently. Additional "Empty" cycle was taken to image the autofluorescence of the human brain tissue. After the last imaging cycle, we added 80% formamide in 2x SSC buffer to remove RNAscope probes completely and stained the nuclei with 5 μ M DRAQ5 at room temperature for 10 min.

Spot decoding

We directly leveraged the computational pipeline in DART-FISH data processing to decode spots in RNAscope images. We registered the z-stack images, max-intensity projected the images, and subtracted the autofluorescence imaged in the "Empty" cycle. Then, the RNAscope spots were decoded by SparseDevolution-Decoder.

Cell annotation

We used anndata[174–176] (v0.8.0) and scanpy[174](v1.9.1) to handle and analyze the data. The data normalization was performed using analytic Pearson residuals[177] (clipped at 40) with a lower bound placed on gene-level standard deviations[178]. Clustering was done with the Leiden algorithm[179] implemented in scanpy.

Annotating the brain short-gene dataset

Cells with counts less than 5 and more than 300 were removed (2,980 out of 26,348). The top 100 highly variable genes (`scanpy.experimental.pp.highly_variable_gene(., flavor='pearson_residuals')`) were used for normalization, embedding and annotations. PCA was performed on Pearson residuals, and the neighborhood graph was created with this command `scanpy.pp.neighbors(n_neighbors=20, n_pcs=15, met-`

ric='cosine'). Single-nucleus RNA-seq reference from Jorstad et al.[126] was subsetted to M1C cells and normalized in the same way as DART-FISH. Pax6 and Scng subclasses were removed since we did not design our probe set to target those subclasses. Average normalized counts (centroids) were computed for every other subclass in the "within_area_subclass" slot and all clusters of DART-FISH. To annotate the DART-FISH clusters at the class level (excitatory, inhibitory, non-neuronal), we first correlated each cluster to all single-nucleus subclasses, and assigned that cluster to the class of the most highly correlated subclass. Annotation of each class was done separately. For excitatory neurons, all DART-FISH cells that had a class label of "excitatory" and had at least 20 transcripts were kept (5,957 cells). We realized that the Leiden clustering was unstable and by mere shuffling of the order of cells, we would obtain very different clusters. We reasoned that by removing some cells that tend to move between clusters, we could get more stable clusters and have more confidence in their annotation. To find cells that don't stably cluster, we ran clustering 20 times, every time shuffling the order of the cells. For every cell, we calculated the number of times it was co-clustered with every other cell and took the average of the non-zero values as the co-clustering index (CCI). A perfect CCI of 20 means that the cell is clustered with the same partners in every clustering instance, while lower values show deviations from this limit. We removed the cells with a CCI smaller than 6 and repeated this filtering procedure for three more iterations. The final results show a more stable clustering of the remaining 5101 cells. We then constructed a new neighborhood graph using newly computed principal components (`n_neighbors=10`, `n_pcs=15`), followed by Leiden clustering. The cluster centroids were calculated and correlated to the reference subclass centroids. We assigned clusters to their maximally correlated reference subclass if we could also see differential expression of their marker genes (scanpy's `rank_genes_groups`), otherwise we labeled them as "NA". Of note, the DART-FISH population labeled as L6b/CT was highly correlated with reference subclasses L6b and L6 CT (Supplementary Figure S5b) and showed expression of marker genes from both subclasses. For inhibitory neurons and non-neuronal

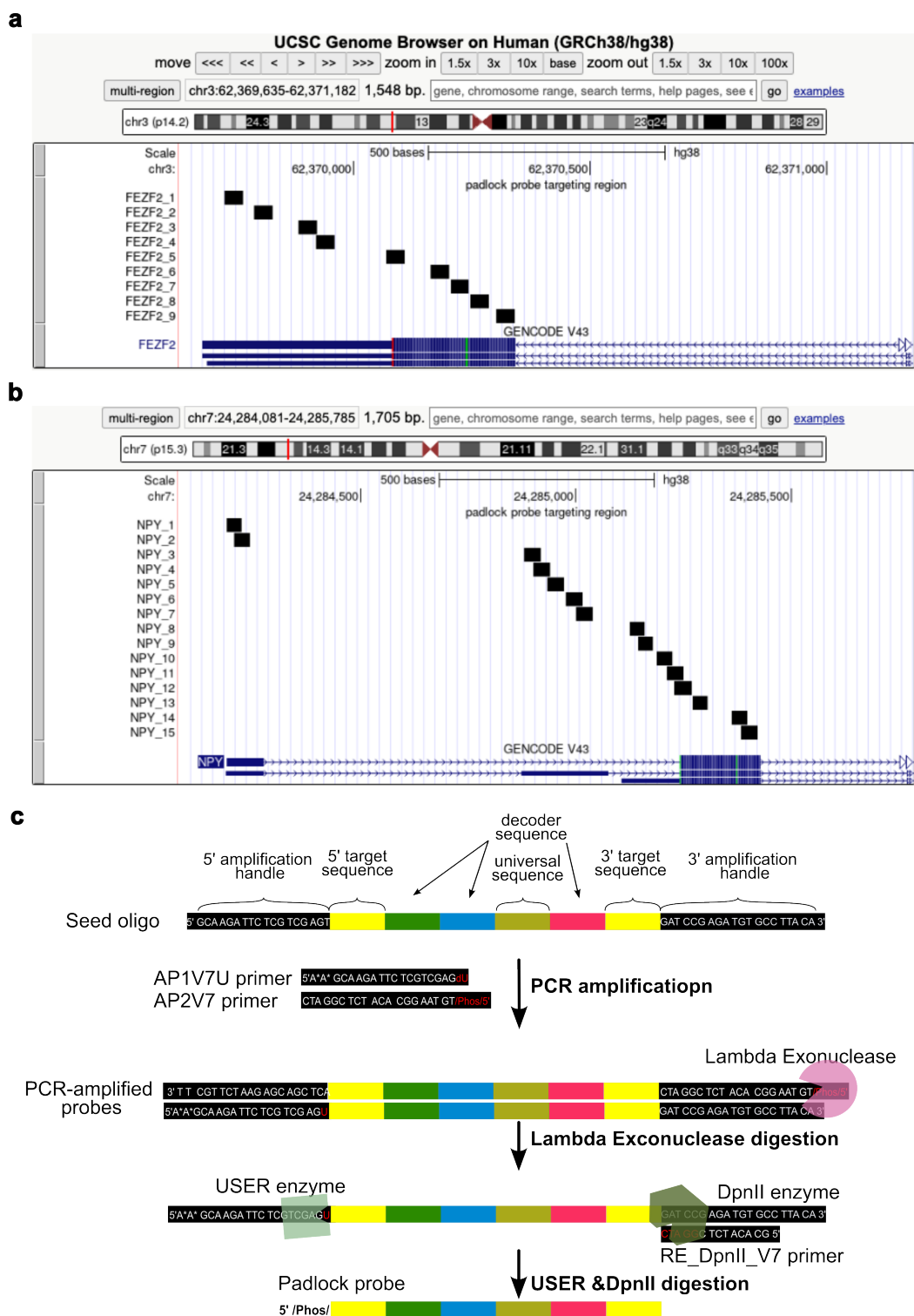
cells, the clustering was more stable to begin with, and we started by constructing the neighborhood matrix (for inhibitory neurons: `n_neighbors=20`, `n_pcs=10`. For non-neuronal cells: `n_neighbors=25`, `n_pcs=15`) and clustering. Then clusters were assigned to the reference subclass with maximum Pearson’s correlation if the marker genes matched, or otherwise were labeled as NA.

Drawing cortical layer boundaries

Cortical layer boundaries were automatically drawn via Support Vector Machine (SVM) decision boundaries. The Scikit-learn python package (v1.1.3) was used to train a SVM on the following excitatory neuron subtype labels: "L2/3 IT", "L4 IT", "L5 IT", "L6 IT", "L6b/CT". First, cells with fewer than 10 total gene counts were filtered out. The x and y coordinates of the cells were standardized via the `StandardScaler()` function, and the data is fed into a SVM with a radial basis function (RBF) kernel with balanced class weights and one vs. one decision function. The RBF SVM model was then applied to a meshgrid with a fine step size with the same geometric size as the original tissue image. The trained SVM classified the cell type label of each point on the meshgrid to define borders between the cortical layers specified by the excitatory neuron subclasses. We drew contours based on the borders between the various subclasses, and manually superimposed them onto Figure 2.3c.

SUPPLEMENTAL FIGURES

Figure S1 **Probe production strategy** (a) A genome browser view showing the target locations for individual padlock probes. Up to 50 padlock probes are designed to tile to constitutive exons. (b) To obtain more probes targeting short genes, we allow overlapping target sequences for padlock probes (*NPY* as an example with 893 nucleotides) (c) Enzymatic production of padlock probes from an oligo pool. A probe set is PCR amplified with a pair of probe-set specific amplification primers from the oligo pool. The forward primer carries a 5' phosphorothioate modification to prevent digestion, while the reverse primer carries a 5' phosphorylation modification to promote digestion by lambda exonuclease. The 5' and 3' ends of the amplicons are cleaved by USER and DpnII enzyme, respectively, to obtain single-stranded padlock probes with 5' phosphate.



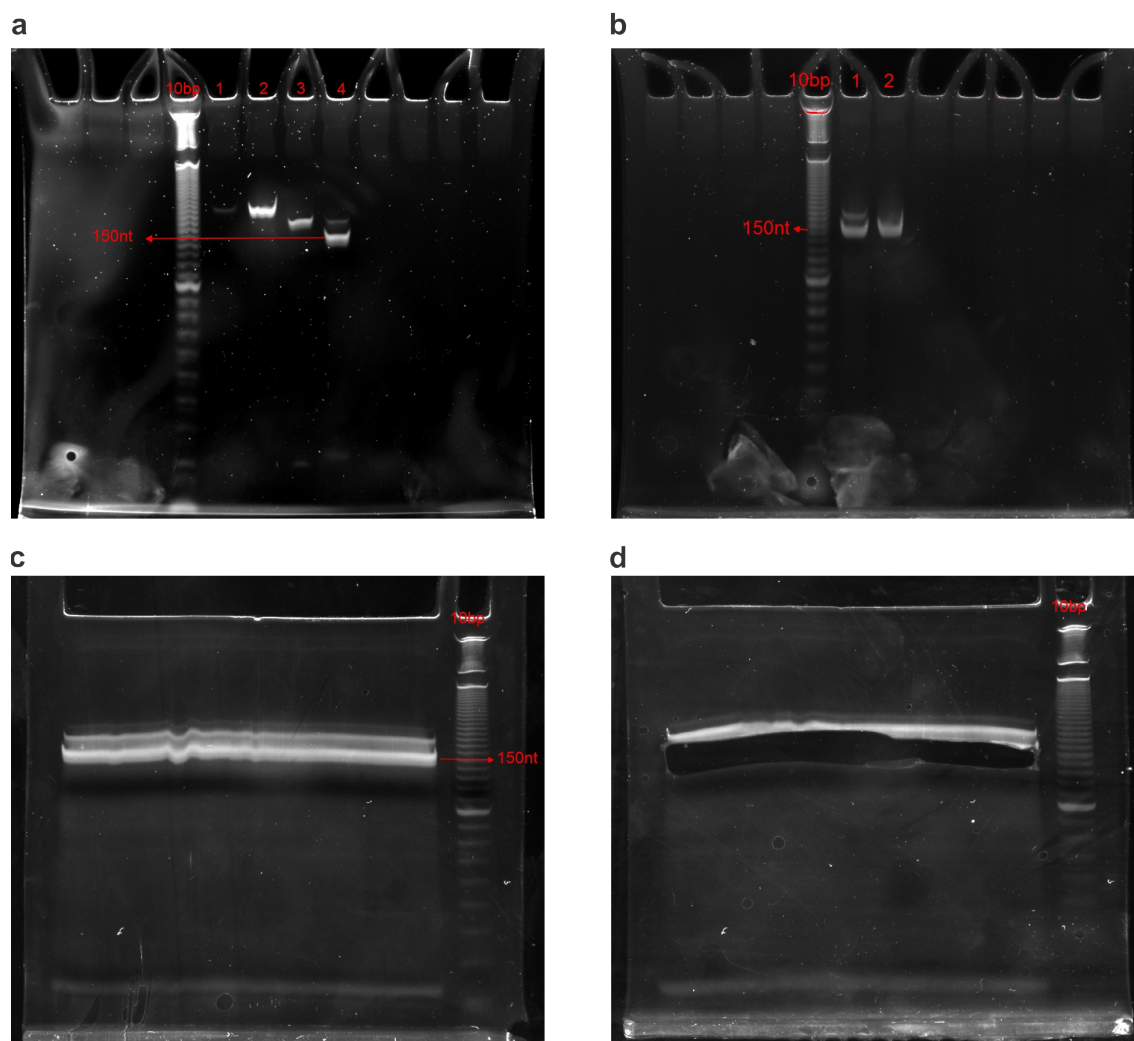
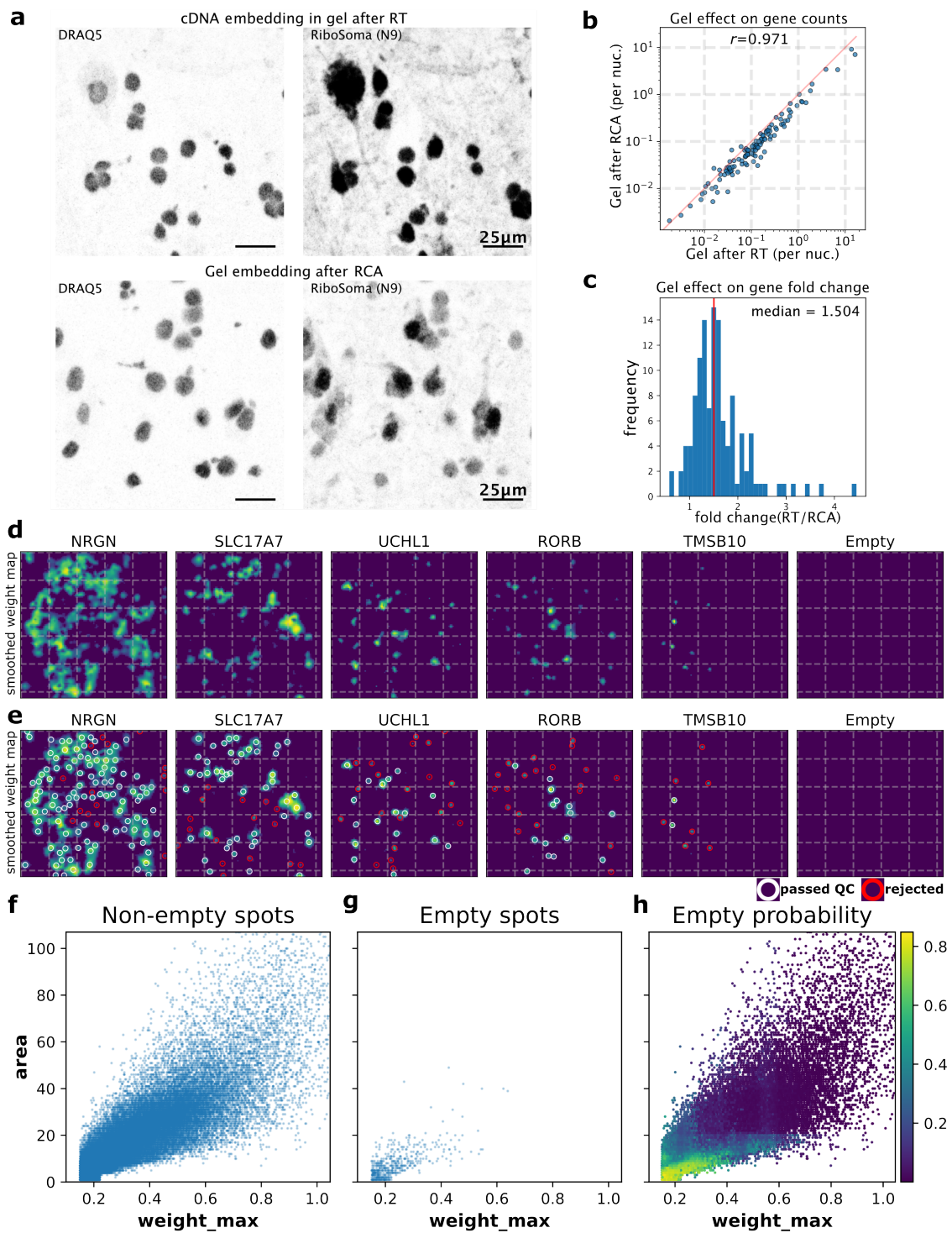


Figure S2 TBU gel check and size selection of padlock probes (a) Small quantities of PCR-amplified and enzyme digested probes were added to TBU gel. Run the gel electrophoresis to check size of each product (10bp: 10bp ladder; 1: PCR-amplified; 2: Lambda Exo digested; 3: Lambda Exo and USER enzyme digested; 4: Lambda Exo, USER, and DpnII enzyme digested). (b) TBU gel check of padlock probes before and after size selection (1: probes before size selection (c); 2: after size selection (cut-out gel in (d))). (c) TBU gel electrophoresis of enzyme digested padlock probes. (d) Size of padlock probes at 150nt were cut out for purification.

Figure S3 Gel embedding after reverse transcription, spot calling, and empty probabilities (a) Fluorescent signal from RiboSoma (randomly primed cDNA) is stronger when the cDNA is embedded in a polyacrylamide gel immediately after reverse transcription (top) compared to gel embedding after rolling circle amplification (bottom). The nuclear DRAQ5 stain does not show a difference between the two conditions (left). RT: reverse-transcription, RCA: rolling circle amplification. (b) Scatter plot comparing the average gene count per nucleus in the two conditions in (a). Each dot represents a gene. Nuclear segmentation was chosen for counting cells to provide a fair comparison between the two conditions. (c) Histogram of the gene expression fold change in (b). 50% of cells show at least a 50% increase in their counts. (d) Smoothed OLS map from **Fig. 1f**. Lasso weight maps (**Fig. 1f**) undergo pixel-wise elbow filtering to select the top 1 or 2 barcodes per pixel. Unbiased weights are then obtained by fitting an ordinary linear regression (OLS) using the selected barcodes (OLS maps). OLS maps are then smoothed with a Gaussian low pass filter. (e) Spot detection on weight maps. For each gene, the local peaks are detected on the respective smoothed OLS map. These peaks then serve as markers for watershed segmentation. The centroids of the segmented areas are used as spot coordinates. White and red circles are drawn around high quality and rejected spots, respectively. (f-g) Scatter plots of two main features extracted from segmented spots with a used barcode (f) and empty (unused) barcodes (g). Empty barcodes tend to be smaller in area and have lower weights than non-empty ones. (h) Emptiness probabilities inferred from a random forest that was trained to distinguish empty from non-empty spots based on the extracted features (`weight_max`, `weight_mean`, `area`). A cutoff is later set on the empty probabilities to keep high quality spots.



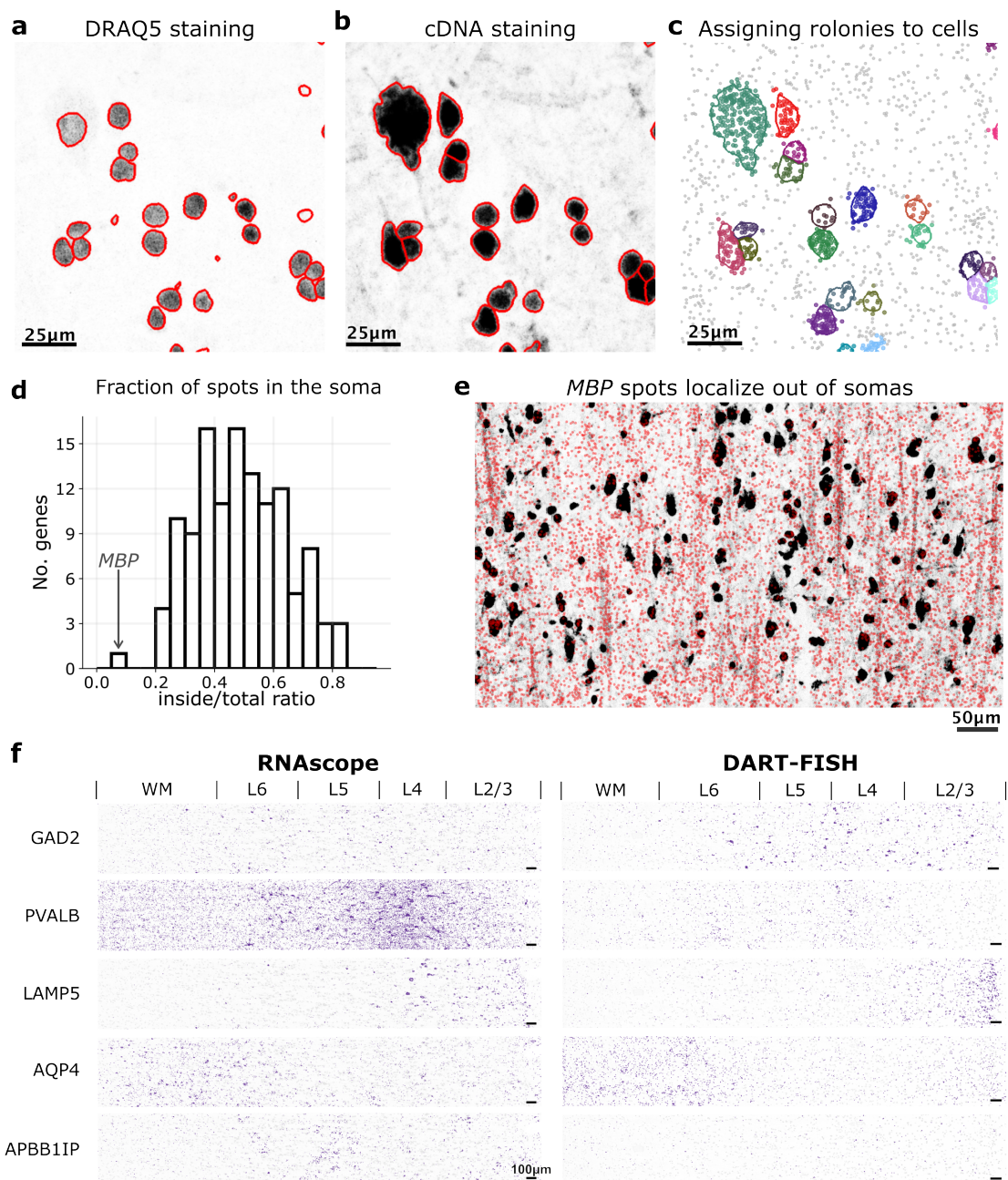


Figure S4 RiboSoma staining, spot assignment, and RNAscope validation (a) Segmentation of cells using only DRAQ5 (nuclear) or (b) using both DRAQ5 and RiboSoma (randomly primed) staining. RiboSoma resolves cell bodies more confidently. (c) Assigning decoded colonies to the closest segmented cell. Transcripts that are too far from cell boundaries are discarded. (d) Histogram showing the fraction of spots inside the segmented cells for each gene. *MBP* encoding myelin base protein, has the lowest fraction of spots inside the cells. (e) An example of MBP being expressed outside the soma. Every red dot is a decoded MBP transcript on the background of RiboSoma (contrast is increased for clarity). MBP spots seem to co-localize with the RiboSoma signal over long threads that resemble axons. (f) RNAscope validation of DART-FISH in the human M1C with genes *GAD2*, *PVALB*, *LAMP5*, *AQP4*, *APBB1IP*.

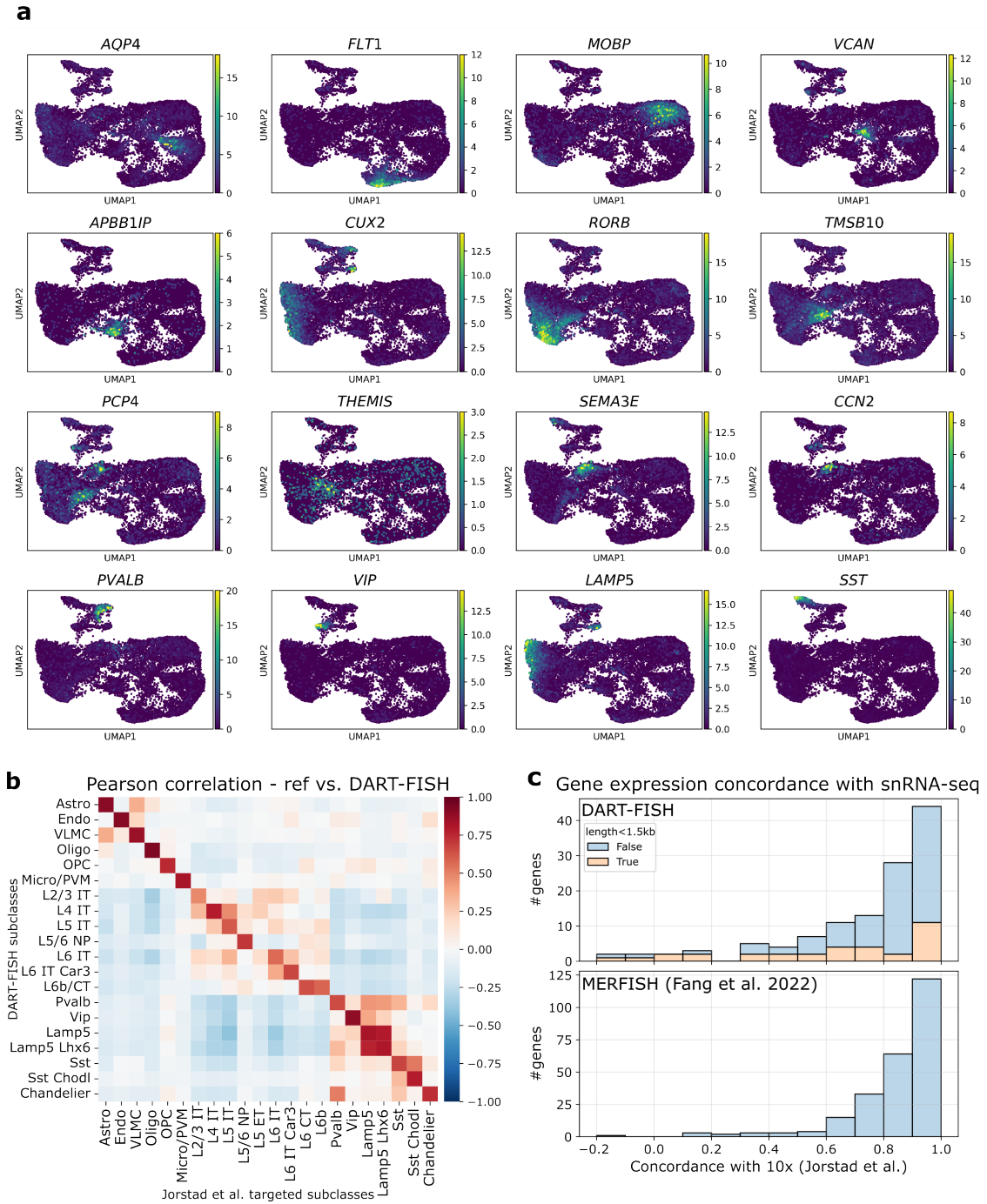


Figure S5 Cell annotation of human M1C (a) UMAP of the DART-FISH M1C data colored by transcripts counts for 16 cell type markers (b) Pearson's correlation of DART-FISH subclasses with the snRNA-seq reference subclasses used for annotation[126] (c) Histogram of concordances values for genes in DART-FISH (top) and MERFISH (bottom, sample H18.06.006.MTG.250.expand.rep1[124]). Concordance is defined as the Pearson's correlation of expression levels across subclasses between SNARE-seq2[113] and the spatial assay. The histogram for DART-FISH is color coded to show the performance of short genes (exon length < 1.5kb)

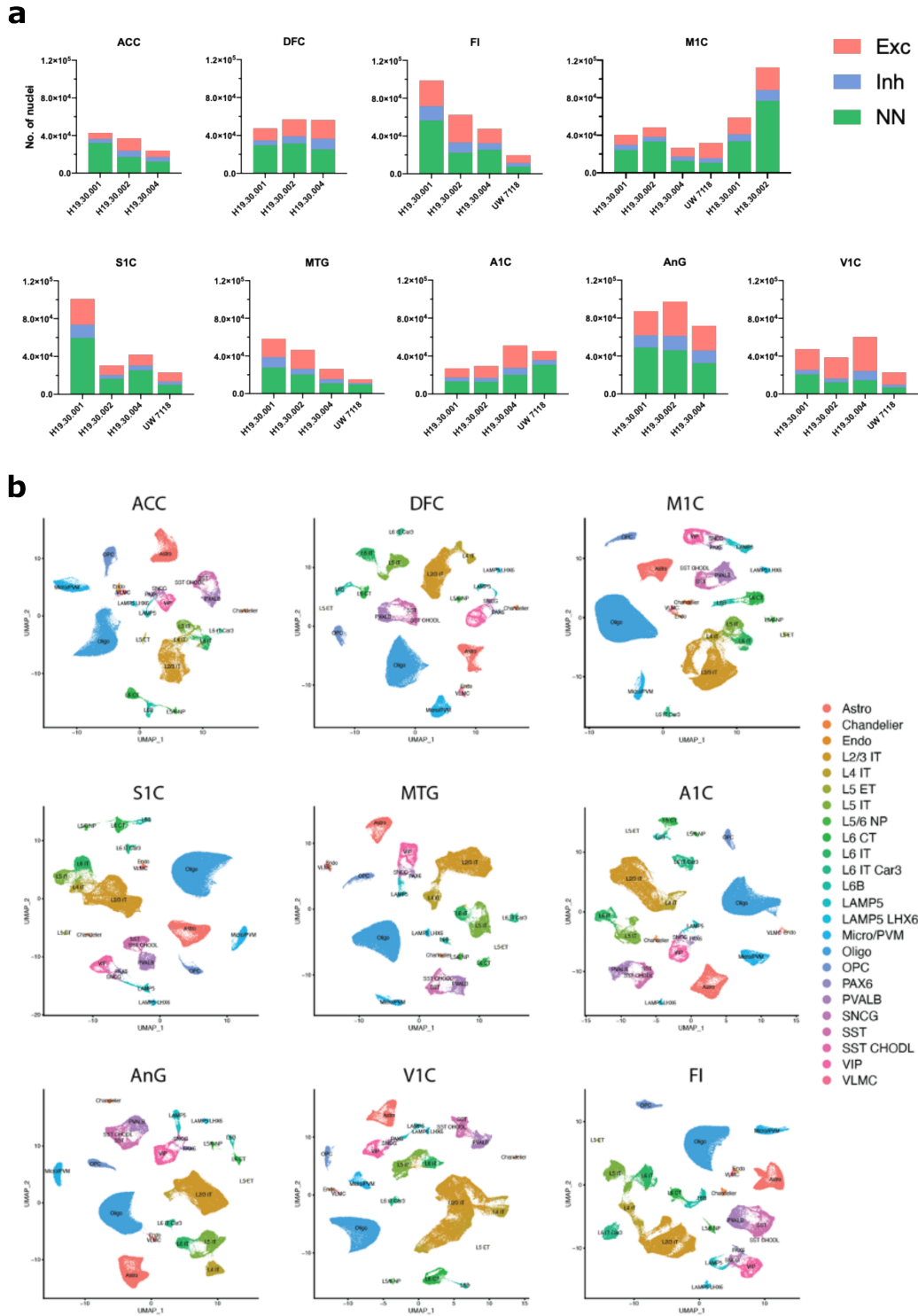
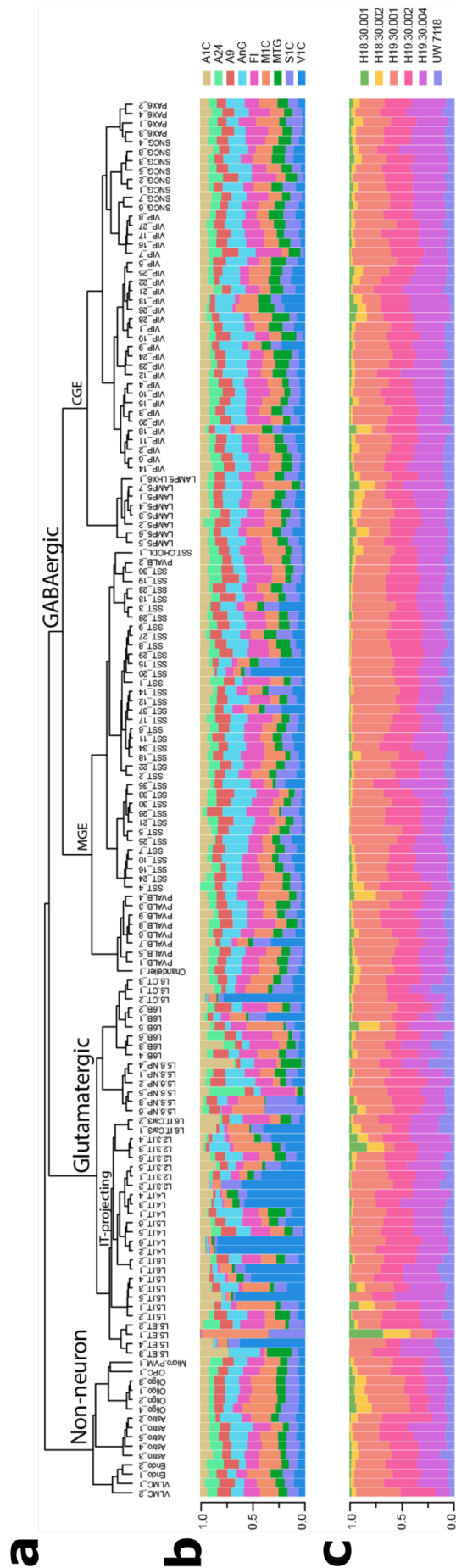


Figure S6 Summary and annotation of SNARE-seq nuclei across 9 areas of human cerebral cortex (a) Barplots of SNARE-seq nuclei assayed from ACC, DFC, FI, M1C, S1C, MTG, A1C, AnG, and V1C areas of the cerebral cortex from multiple human brains. The x labels represent the human brain IDs. Exc: Excitatory neurons, Inh: Inhibitory neurons, NN: Non-neuronal cells (b) UMAP plots of 24 annotated excitatory neurons (L2/3 IT, L4 IT, L5 IT, L5 ET, L5/6 NP, L6 IT, L6 IT Car3, L6 CT, and L6B), inhibitory neurons (PVALB, Chandelier, SST, SST CHODL, LAMP5 LHX6, LAMP5, SNCG, VIP, and PAX6), and non-neuronal (Astro, Endo, Micro/PVM, Oligo, OPC, VLNC) subclasses in each cortical area.



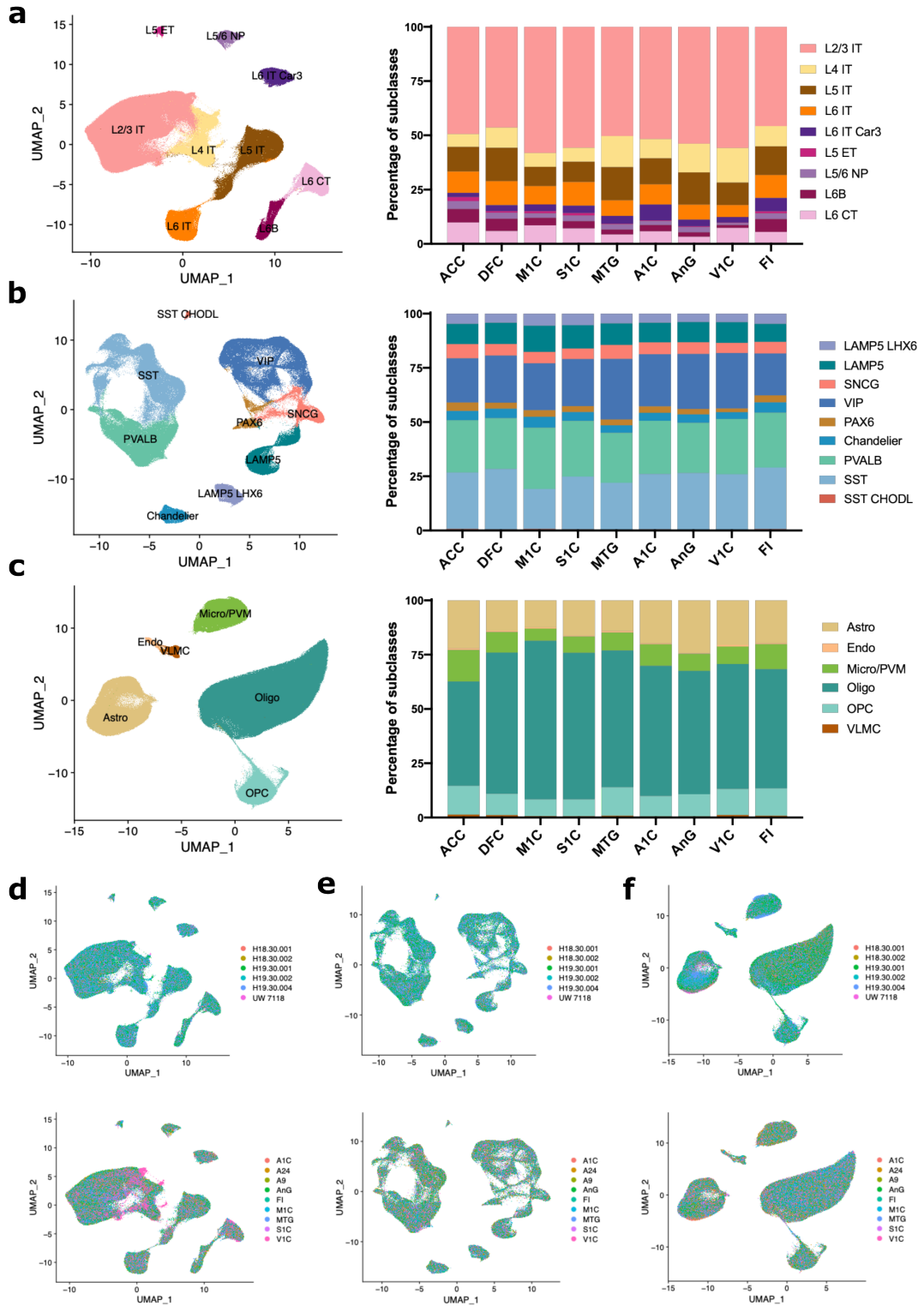


Figure S8 Composition of 24 subclasses across 9 areas of human cerebral cortex
The excitatory neurons, inhibitory neurons, and non-neuronal cells from multiple human brains were pooled together and clustered separately. UMAP plots and composition of each subclass of excitatory neurons, inhibitory neurons, and non-neuronal cells across 9 cortical areas are shown in (a), (b), and (c), respectively. UMAP plots in (d), (e), and (f) represent the excitatory neurons, inhibitory neurons, and non-neuronal cells labeled by the human brain IDs and cortical areas, respectively, in the upper and lower panel.

REFERENCES

1. Herculano-Houzel, S. The human brain in numbers: a linearly scaled-up primate brain. en. *Front. Hum. Neurosci.* **3**, 31 (Nov. 2009).
2. Noback, C. R., Ruggiero, D. A., Strominger, N. L. & Demarest, R. J. *The Human Nervous System: Structure and Function* en (Springer Science & Business Media, 2005).
3. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. en. *Nat. Rev. Neurosci.* **10**, 724–735 (Oct. 2009).
4. Carroll, S. B. Genetics and the making of Homo sapiens. en. *Nature* **422**, 849–857 (Apr. 2003).
5. Cáceres, M., Lachuer, J., Zapala, M. A., Redmond, J. C., Kudo, L., Geschwind, D. H., Lockhart, D. J., Preuss, T. M. & Barlow, C. Elevated gene expression levels distinguish human from non-human primate brains. en. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13030–13035 (Oct. 2003).
6. Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues von Dr. K. Brodmann* de (J.A. Barth, 1909).
7. Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L. & Petersen, S. E. Functional network organization of the human brain. en. *Neuron* **72**, 665–678 (Nov. 2011).
8. Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H. & Buckner, R. L. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (Sept. 2011).

9. Rubenstein, J. L., Shimamura, K., Martinez, S. & Puelles, L. Regionalization of the prosencephalic neural plate. en. *Annu. Rev. Neurosci.* **21**, 445–477 (1998).
10. Mallamaci, A. & Stoykova, A. Gene networks controlling early cerebral cortex arealization. en. *Eur. J. Neurosci.* **23**, 847–856 (Feb. 2006).
11. O’Leary, D. D. M., Chou, S.-J. & Sahara, S. Area patterning of the mammalian cortex. en. *Neuron* **56**, 252–269 (Oct. 2007).
12. Bayatti, N., Sarma, S., Shaw, C., Eyre, J. A., Vouyiouklis, D. A., Lindsay, S. & Clowry, G. J. Progressive loss of PAX6, TBR2, NEUROD and TBR1 mRNA gradients correlates with translocation of EMX2 to the cortical plate during human cortical development. en. *Eur. J. Neurosci.* **28**, 1449–1456 (Oct. 2008).
13. Sidman, R. L. & Rakic, P. Neuronal migration, with special reference to developing human brain: a review. en. *Brain Res.* **62**, 1–35 (Nov. 1973).
14. Nowakowski, T. J., Pollen, A. A., Sandoval-Espinosa, C. & Kriegstein, A. R. Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. en. *Neuron* **91**, 1219–1227 (Sept. 2016).
15. Kriegstein, A., Noctor, S. & Martinez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. en. *Nat. Rev. Neurosci.* **7**, 883–890 (Nov. 2006).
16. Lukasiewicz, A., Savatier, P., Cortay, V., Giroud, P., Huissoud, C., Berland, M., Kennedy, H. & Dehay, C. G1 phase regulation, area-specific cell cycle control, and cytoarchitectonics in the primate cortex. en. *Neuron* **47**, 353–364 (Aug. 2005).
17. LaMonica, B. E., Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Mitotic spindle orientation predicts outer radial glial cell generation in human neocortex. en. *Nat. Commun.* **4**, 1665 (2013).
18. Rakic, P. Neuronal migration and contact guidance in the primate telencephalon. en. *Postgrad. Med. J.* **54 Suppl 1**, 25–40 (1978).

19. Rakic, P. Radial unit hypothesis of neocortical expansion. en. *Novartis Found. Symp.* **228**, 30–42, discussion 42–52 (2000).
20. Noctor, S. C., Martinez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. en. *Nat. Neurosci.* **7**, 136–144 (Feb. 2004).
21. Betizeau, M., Cortay, V., Patti, D., Pfister, S., Gautier, E., Bellemin-Ménard, A., Afanassieff, M., Huissoud, C., Douglas, R. J., Kennedy, H. & Dehay, C. Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. en. *Neuron* **80**, 442–457 (Oct. 2013).
22. Malik, S., Vinukonda, G., Vose, L. R., Diamond, D., Bhimavarapu, B. B. R., Hu, F., Zia, M. T., Hevner, R., Zecevic, N. & Ballabh, P. Neurogenesis continues in the third trimester of pregnancy and is suppressed by premature birth. en. *J. Neurosci.* **33**, 411–423 (Jan. 2013).
23. Tan, X. & Shi, S.-H. Neocortical neurogenesis and neuronal migration. en. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 443–459 (July 2013).
24. Marín-Padilla, M. Cajal–Retzius cells and the development of the neocortex. en. *Trends Neurosci.* **21**, 64–71 (Feb. 1998).
25. Causeret, F., Moreau, M. X., Pierani, A. & Blanquie, O. The multiple facets of Cajal–Retzius neurons. en. *Development* **148** (June 2021).
26. Ohtaka-Maruyama, C. Subplate Neurons as an Organizer of Mammalian Neocortical Development. en. *Front. Neuroanat.* **14**, 8 (Mar. 2020).
27. Angevine Jr, J. B. & Sidman, R. L. Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. en. *Nature* **192**, 766–768 (Nov. 1961).
28. Harris, K. D. & Shepherd, G. M. G. The neocortical circuit: themes and variations. en. *Nat. Neurosci.* **18**, 170–181 (Feb. 2015).

29. Zeng, H., Shen, E. H., Hohmann, J. G., Oh, S. W., Bernard, A., Royall, J. J., Glattfelder, K. J., Sunkin, S. M., Morris, J. A., Guillozet-Bongaarts, A. L., Smith, K. A., Ebbert, A. J., Swanson, B., Kuan, L., Page, D. T., Overly, C. C., Lein, E. S., Hawrylycz, M. J., Hof, P. R., Hyde, T. M., Kleinman, J. E. & Jones, A. R. Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. en. *Cell* **149**, 483–496 (Apr. 2012).
30. Hodge, R. D., Bakken, T. E., Miller, J. A., Smith, K. A., Barkan, E. R., Graybuck, L. T., Close, J. L., Long, B., Johansen, N., Penn, O., Yao, Z., Eggermont, J., Höllt, T., Levi, B. P., Shehata, S. I., Aeversmann, B., Beller, A., Bertagnolli, D., Brouner, K., Casper, T., Cobbs, C., Dalley, R., Dee, N., Ding, S.-L., Ellenbogen, R. G., Fong, O., Garren, E., Goldy, J., Gwinn, R. P., Hirschstein, D., Keene, C. D., Keshk, M., Ko, A. L., Lathia, K., Mahfouz, A., Maltzer, Z., McGraw, M., Nguyen, T. N., Nyhus, J., Ojemann, J. G., Oldre, A., Parry, S., Reynolds, S., Rimorin, C., Shapovalova, N. V., Somasundaram, S., Szafer, A., Thomsen, E. R., Tieu, M., Quon, G., Scheuermann, R. H., Yuste, R., Sunkin, S. M., Lelieveldt, B., Feng, D., Ng, L., Bernard, A., Hawrylycz, M., Phillips, J. W., Tasic, B., Zeng, H., Jones, A. R., Koch, C. & Lein, E. S. Conserved cell types with divergent features in human versus mouse cortex. en. *Nature* **573**, 61–68 (Sept. 2019).
31. Wichterle, H., Garcia-Verdugo, J. M., Herrera, D. G. & Alvarez-Buylla, A. Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. en. *Nat. Neurosci.* **2**, 461–466 (May 1999).
32. Nery, S., Fishell, G. & Corbin, J. G. The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. en. *Nat. Neurosci.* **5**, 1279–1287 (Dec. 2002).
33. Gelman, D. M., Martini, F. J., Nóbrega-Pereira, S., Pierani, A., Kessaris, N. & Marin, O. The embryonic preoptic area is a novel source of cortical GABAergic interneurons. en. *J. Neurosci.* **29**, 9380–9389 (July 2009).

34. Gelman, D. M., Marm, O. & Rubenstein, J. L. R. in *Jasper's Basic Mechanisms of the Epilepsies* (eds Noebels, J. L., Avoli, M., Rogawski, M. A., Olsen, R. W. & Delgado-Escueta, A. V.) (National Center for Biotechnology Information (US), Bethesda (MD)).
35. Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G. & Wu, C. Interneurons of the neocortical inhibitory system. en. *Nat. Rev. Neurosci.* **5**, 793–807 (Oct. 2004).
36. Harwell, C. C., Fuentealba, L. C., Gonzalez-Cerrillo, A., Parker, P. R. L., Gertz, C. C., Mazzola, E., Garcia, M. T., Alvarez-Buylla, A., Cepko, C. L. & Kriegstein, A. R. Wide Dispersion and Diversity of Clonally Related Inhibitory Interneurons. en. *Neuron* **87**, 999–1007 (Sept. 2015).
37. Lodato, S., Rouaux, C., Quast, K. B., Jantrachotechatchawan, C., Studer, M., Hensch, T. K. & Arlotta, P. Excitatory projection neuron subtypes control the distribution of local inhibitory interneurons in the cerebral cortex. en. *Neuron* **69**, 763–779 (Feb. 2011).
38. Lee, A. T., Gee, S. M., Vogt, D., Patel, T., Rubenstein, J. L. & Sohal, V. S. Pyramidal neurons in prefrontal cortex receive subtype-specific forms of excitation and inhibition. en. *Neuron* **81**, 61–68 (Jan. 2014).
39. Von Economo, C. F. & Koskinas, G. N. *Die Cytoarchitektonik der Hirnrinde des erwachsenen Menschen* de (J. Springer, 1925).
40. Charvet, C. J., Cahalane, D. J. & Finlay, B. L. Systematic, cross-cortex variation in neuron numbers in rodents and primates. en. *Cereb. Cortex* **25**, 147–160 (Jan. 2015).
41. Rakic, P. Specification of cerebral cortical areas. en. *Science* **241**, 170–176 (July 1988).
42. Creutzfeldt, O. D. Generality of the functional structure of the neocortex. en. *Naturwissenschaften* **64**, 507–517 (Oct. 1977).

43. Van der Loos, H. & Woolsey, T. A. Somatosensory cortex: structural alterations following early injury to sense organs. en. *Science* **179**, 395–398 (Jan. 1973).
44. O’Leary, D. D. Do cortical areas emerge from a protocortex? en. *Trends Neurosci.* **12**, 400–406 (Oct. 1989).
45. Nano, P. R., Nguyen, C. V., Mil, J. & Bhaduri, A. Cortical Cartography: Mapping Arealization Using Single-Cell Omics Technology. en. *Front. Neural Circuits* **15**, 788560 (Dec. 2021).
46. Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G. & Nowakowski, T. J. Development and Arealization of the Cerebral Cortex. en. *Neuron* **103**, 980–1004 (Sept. 2019).
47. Bhaduri, A., Sandoval-Espinosa, C., Otero-Garcia, M., Oh, I., Yin, R., Eze, U. C., Nowakowski, T. J. & Kriegstein, A. R. An atlas of cortical arealization identifies dynamic molecular signatures. en. *Nature* **598**, 200–204 (Oct. 2021).
48. Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshev, D., Ounadjela, J. R., Shuga, J., Wang, X., Lim, D. A., West, J. A., Leyrat, A. A., Kent, W. J. & Kriegstein, A. R. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. en. *Science* **358**, 1318–1323 (Dec. 2017).
49. De Marco Garcia, N. V., Priya, R., Tuncdemir, S. N., Fishell, G. & Karayannis, T. Sensory inputs control the integration of neurogliaform interneurons into cortical circuits. en. *Nat. Neurosci.* **18**, 393–401 (Mar. 2015).
50. Tuncdemir, S. N., Wamsley, B., Stam, F. J., Osakada, F., Goulding, M., Callaway, E. M., Rudy, B. & Fishell, G. Early Somatostatin Interneuron Connectivity Mediates the Maturation of Deep Layer Cortical Circuits. en. *Neuron* **89**, 521–535 (Feb. 2016).

51. Larsen, R., Proue, A., Scott, E. P., Christiansen, M. & Nakagawa, Y. The Thalamus Regulates Retinoic Acid Signaling and Development of Parvalbumin Interneurons in Postnatal Mouse Prefrontal Cortex. en. *eNeuro* **6** (Mar. 2019).
52. Chou, S.-J., Babot, Z., Leingärtner, A., Studer, M., Nakagawa, Y. & O’Leary, D. D. M. Geniculocortical input drives genetic distinctions between primary and higher-order visual areas. en. *Science* **340**, 1239–1242 (June 2013).
53. Kang, H. J., Kawasaki, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A. M. M., Pletikos, M., Meyer, K. A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M. B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S. N., Vortmeyer, A., Weinberger, D. R., Mane, S., Hyde, T. M., Huttner, A., Reimers, M., Kleinman, J. E. & Sestan, N. Spatio-temporal transcriptome of the human brain. en. *Nature* **478**, 483–489 (Oct. 2011).
54. Pletikos, M., Sousa, A. M. M., Sedmak, G., Meyer, K. A., Zhu, Y., Cheng, F., Li, M., Kawasaki, Y. I. & Sestan, N. Temporal specification and bilaterality of human neocortical topographic gene expression. en. *Neuron* **81**, 321–332 (Jan. 2014).
55. Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Chakravarty, M. M., Chapin, M., Chong, J., Dalley, R. A., David Daly, B., Dang, C., Datta, S., Dee, N., Dolbeare, T. A., Faber, V., Feng, D., Fowler, D. R., Goldy, J., Gregor, B. W., Haradon, Z., Haynor, D. R., Hohmann, J. G., Horvath, S., Howard, R. E., Jeromin, A., Jochim, J. M., Kinnunen, M., Lau, C., Lazarz, E. T., Lee, C., Lemon, T. A., Li, L., Li, Y., Morris, J. A., Overly, C. C., Parker, P. D., Parry, S. E., Reding, M., Royall, J. J., Schulkin, J., Sequeira, P. A., Slaughterbeck, C. R., Smith, S. C., Sodt, A. J., Sunkin, S. M., Swanson, B. E., Vawter, M. P., Williams, D., Wohnoutka, P., Zielke, H. R., Geschwind, D. H., Hof, P. R., Smith, S. M., Koch, C., Grant, S. G. N.

- & Jones, A. R. An anatomically comprehensive atlas of the adult human brain transcriptome. en. *Nature* **489**, 391–399 (Sept. 2012).
56. Xu, C., Li, Q., Efimova, O., He, L., Tatsumoto, S., Stepanova, V., Oishi, T., Udonon, T., Yamaguchi, K., Shigenobu, S., Kakita, A., Nawa, H., Khaitovich, P. & Go, Y. Human-specific features of spatial gene expression and regulation in eight brain regions. en. *Genome Res.* **28**, 1097–1110 (Aug. 2018).
 57. Zhu, Y., Sousa, A. M. M., Gao, T., Skarica, M., Li, M., Santpere, G., Esteller-Cucala, P., Juan, D., Ferrández-Peral, L., Gulden, F. O., Yang, M., Miller, D. J., Marques-Bonet, T., Imamura Kawasawa, Y., Zhao, H. & Sestan, N. Spatiotemporal transcriptomic divergence across human and macaque brain development. en. *Science* **362** (Dec. 2018).
 58. Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O. V., Gulden, F. O., Pochareddy, S., Sunkin, S. M., Li, Z., Shin, Y., Zhu, Y., Sousa, A. M. M., Werling, D. M., Kitchen, R. R., Kang, H. J., Pletikos, M., Choi, J., Muchnik, S., Xu, X., Wang, D., Lorente-Galdos, B., Liu, S., Giusti-Rodriguez, P., Won, H., de Leeuw, C. A., Pardiñas, A. F., BrainSpan Consortium, PsychENCODE Consortium, PsychENCODE Developmental Subgroup, Hu, M., Jin, F., Li, Y., Owen, M. J., O’Donovan, M. C., Walters, J. T. R., Posthuma, D., Reimers, M. A., Levitt, P., Weinberger, D. R., Hyde, T. M., Kleinman, J. E., Geschwind, D. H., Hawrylycz, M. J., State, M. W., Sanders, S. J., Sullivan, P. F., Gerstein, M. B., Lein, E. S., Knowles, J. A. & Sestan, N. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. en. *Science* **362** (Dec. 2018).
 59. Yao, Z., van Velthoven, C. T. J., Nguyen, T. N., Goldy, J., Sedenon-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S.-L., Fong, O., Garren, E., Glandon, A., Gouwens, N. W., Gray, J., Graybuck, L. T., Hawrylycz, M. J., Hirschstein, D., Kroll, M., Lathia, K., Lee, C., Levi, B., McMillen, D., Mok, S., Pham, T., Ren, Q., Rimorin, C., Shapovalova, N., Sulc, J., Sunkin, S. M., Tieu, M., Torkelson, A., Tung, H., Ward, K., Dee, N.,

- Smith, K. A., Tasic, B. & Zeng, H. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. en. *Cell* **184**, 3222–3241.e26 (June 2021).
60. Zeng, H. What is a cell type and how to define it? en. *Cell* **185**, 2739–2755 (July 2022).
61. Ramón y Cajal, S. *Histologie du système nerveux de l’homme & des vertébrés: Cervelet, cerveau moyen, rétine, couche optique, corps strié, écorce cérébrale générale & régionale, grand sympathique* fr (A. Maloine, 1911).
62. Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K. E., Lathia, K., Rimorin, C., Tieu, M., Larsen, R., Casper, T., Barkan, E., Kroll, M., Parry, S., Shapovalova, N. V., Hirschstein, D., Pendergraft, J., Sullivan, H. A., Kim, T. K., Szafer, A., Dee, N., Groblewski, P., Wickersham, I., Cetin, A., Harris, J. A., Levi, B. P., Sunkin, S. M., Madisen, L., Daigle, T. L., Looger, L., Bernard, A., Phillips, J., Lein, E., Hawrylycz, M., Svoboda, K., Jones, A. R., Koch, C. & Zeng, H. Shared and distinct transcriptomic cell types across neocortical areas. en. *Nature* **563**, 72–78 (Nov. 2018).
63. Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P., Poliodakis, D., Regev, A., Sestan, N., Wickersham, I. R. & Zeng, H. The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. en. *Neuron* **96**, 542–557 (Nov. 2017).
64. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. en. *Nature* **598**, 86–102 (Oct. 2021).
65. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. en. *Nat. Rev. Genet.* **16**, 57–66 (Jan. 2015).

66. Close, J. L., Long, B. R. & Zeng, H. Spatially resolved transcriptomics in neuroscience. en. *Nat. Methods* **18**, 23–25 (Jan. 2021).
67. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. en. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (Feb. 2015).
68. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. en. *Nature* **596**, 211–220 (Aug. 2021).
69. Cajigas, I. J., Tushev, G., Will, T. J., tom Dieck, S., Fuerst, N. & Schuman, E. M. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. en. *Neuron* **74**, 453–466 (May 2012).
70. Miyashiro, K., Dichter, M. & Eberwine, J. On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. en. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10800–10804 (Nov. 1994).
71. Poon, M. M., Choi, S.-H., Jamieson, C. A. M., Geschwind, D. H. & Martin, K. C. Identification of process-localized mRNAs from cultured rodent hippocampal neurons. en. *J. Neurosci.* **26**, 13390–13399 (Dec. 2006).
72. Mayford, M., Baranes, D., Podsypanina, K. & Kandel, E. R. The 3-untranslated region of CaMKII α is a cis-acting signal for the localization and translation of mRNA in dendrites. *Proceedings of the National Academy of Sciences* **93**, 13250–13255 (1996).
73. Miller, S., Yasuda, M., Coats, J. K., Jones, Y., Martone, M. E. & Mayford, M. Disruption of dendritic translation of CaMKII α impairs stabilization of synaptic plasticity and memory consolidation. en. *Neuron* **36**, 507–519 (Oct. 2002).
74. Martin, K. C., Barad, M. & Kandel, E. R. Local protein synthesis and its role in synapse-specific plasticity. en. *Curr. Opin. Neurobiol.* **10**, 587–592 (Oct. 2000).
75. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. en. *Nat. Methods* **5**, 877–879 (Oct. 2008).

76. Levesque, M. J. & Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. en. *Nat. Methods* **10**, 246–248 (Mar. 2013).
77. Bhakdi, S. C. & Thaicharoen, P. Easy Employment and Crosstalk-Free Detection of Seven Fluorophores in a Widefield Fluorescence Microscope. en. *Methods Protoc* **1** (June 2018).
78. Codeluppi, S., Borm, L. E., Zeisel, A., La Manno, G., van Lunteren, J. A., Svensson, C. I. & Linnarsson, S. Spatial organization of the somatosensory cortex revealed by osmFISH. en. *Nat. Methods* **15**, 932–935 (Nov. 2018).
79. Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P. I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J. & Frisén, J. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. en. *Science* **353**, 78–82 (July 2016).
80. Rodriques, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., Welch, J., Chen, L. M., Chen, F. & Macosko, E. Z. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. en. *Science* **363**, 1463–1467 (Mar. 2019).
81. Liu, Y., Yang, M., Deng, Y., Su, G., Enninfu, A., Guo, C. C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., Norris, E., Pan, A., Li, J., Xiao, Y., Halene, S. & Fan, R. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. en. *Cell* **183**, 1665–1681.e18 (Dec. 2020).
82. Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., Terry, R., Jeanty, S. S. F., Li, C., Amamoto, R., Peters, D. T., Turczyk, B. M., Marblestone, A. H., Inverso, S. A., Bernard, A., Mali, P., Rios, X., Aach, J. & Church, G. M. Highly multiplexed subcellular RNA sequencing in situ. en. *Science* **343**, 1360–1363 (Mar. 2014).

83. Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Terry, R., Turczyk, B. M., Yang, J. L., Lee, H. S., Aach, J., Zhang, K. & Church, G. M. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. en. *Nat. Protoc.* **10**, 442–458 (Mar. 2015).
84. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. en. *Science* **348**, aaa6090 (Apr. 2015).
85. Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulana, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C. & Cai, L. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. en. *Nature* **568**, 235–239 (Mar. 2019).
86. Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., Nolan, G. P., Bava, F.-A. & Deisseroth, K. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. en. *Science* **361** (July 2018).
87. Gyllborg, D., Langseth, C. M., Qian, X., Choi, E., Salas, S. M., Hilscher, M. M., Lein, E. S. & Nilsson, M. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. en. *Nucleic Acids Res.* **48**, e112 (Nov. 2020).
88. Liu, S., Punthambaker, S., Iyer, E. P. R., Ferrante, T., Goodwin, D., Fürth, D., Pawlowski, A. C., Jindal, K., Tam, J. M., Mifflin, L., Alon, S., Sinha, A., Wassie, A. T., Chen, F., Cheng, A., Willocq, V., Meyer, K., Ling, K.-H., Campisson, C. K., Kohman, R. E., Aach, J., Lee, J. H., Yankner, B. A., Boyden, E. S. & Church, G. M. Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel in situ analyses. en. *Nucleic Acids Res.* **49**, e58 (June 2021).
89. Benavides, S. H., Monserrat, A. J., Fariña, S. & Porta, E. A. Sequential histochemical studies of neuronal lipofuscin in human cerebral cortex from the first to the ninth decade of life. en. *Arch. Gerontol. Geriatr.* **34**, 219–231 (2002).

90. Di Guardo, G. Lipofuscin, lipofuscin-like pigments and autofluorescence. en. *Eur. J. Histochem.* **59**, 2485 (Feb. 2015).
91. Banerjee, B., Miedema, B. E. & Chandrasekhar, H. R. Role of basement membrane collagen and elastin in the autofluorescence spectra of the colon. en. *J. Investig. Med.* **47**, 326–332 (July 1999).
92. Rodrigues, R. M., Macko, P., Palosaari, T. & Whelan, M. P. Autofluorescence microscopy: a non-destructive tool to monitor mitochondrial toxicity. en. *Toxicol. Lett.* **206**, 281–288 (Oct. 2011).
93. Bhargava, P. & Schnellmann, R. G. Mitochondrial energetics in the kidney. en. *Nat. Rev. Nephrol.* **13**, 629–646 (Oct. 2017).
94. Plongthongkum, N., Diep, D., Chen, S., Lake, B. B. & Zhang, K. Scalable dual-omics profiling with single-nucleus chromatin accessibility and mRNA expression sequencing 2 (SNARE-seq2). en. *Nat. Protoc.* **16**, 4992–5029 (Nov. 2021).
95. Femino, A. M. *Visualization of Single RNA Transcripts in Situ* en (University of Massachusetts Graduate School of Biomedical Sciences, 2001).
96. Rodriguez, A. J., Czaplinski, K., Condeelis, J. S. & Singer, R. H. Mechanisms and cellular roles of local protein synthesis in mammalian cells. en. *Curr. Opin. Cell Biol.* **20**, 144–149 (Apr. 2008).
97. Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J. F., Gould, J., Griffin, G. K., Borg, Å., Ronaghi, M., Frisén, J., Lundeberg, J., Regev, A. & Ståhl, P. L. High-definition spatial transcriptomics for in situ tissue profiling. en. *Nat. Methods* **16**, 987–990 (Oct. 2019).
98. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., Wang, X., Lu, H., Chen, X., Liu, X., Huang, X., Li, Z., Hong, Y., Jiang, Y., Peng, J., Liu, S., Shen, M., Liu, C., Li, Q., Yuan, Y., Wei, X., Zheng, H., Feng, W., Wang, Z., Liu, Y., Wang, Z., Yang, Y., Xiang, H., Han, L., Qin,

- B., Guo, P., Lai, G., Muñoz-Cánoves, P., Maxwell, P. H., Thiery, J. P., Wu, Q.-F., Zhao, F., Chen, B., Li, M., Dai, X., Wang, S., Kuang, H., Hui, J., Wang, L., Fei, J.-F., Wang, O., Wei, X., Lu, H., Wang, B., Liu, S., Gu, Y., Ni, M., Zhang, W., Mu, F., Yin, Y., Yang, H., Lisby, M., Cornall, R. J., Mulder, J., Uhlén, M., Esteban, M. A., Li, Y., Liu, L., Xu, X. & Wang, J. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. en. *Cell* **185**, 1777–1792.e21 (May 2022).
99. Fu, X., Sun, L., Dong, R., Chen, J. Y., Silakit, R., Condon, L. F., Lin, Y., Lin, S., Palmiter, R. D. & Gu, L. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. en. *Cell* **185**, 4621–4633.e17 (Nov. 2022).
 100. Moses, L. & Pachter, L. Museum of spatial transcriptomics. en. *Nat. Methods* **19**, 534–546 (May 2022).
 101. Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., Rubinstein, N. D., Hao, J., Regev, A., Dulac, C. & Zhuang, X. Molecular, spatial, and functional single-cell profiling of the hypothalamic pre-optic region. en. *Science* **362** (Nov. 2018).
 102. Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C. & Ward, D. C. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. en. *Nat. Genet.* **19**, 225–232 (July 1998).
 103. Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E. A., Karlin-Neumann, G. A., Fakhrai-Rad, H., Ronaghi, M., Willis, T. D., Landegren, U. & Davis, R. W. Multiplexed genotyping with sequence-tagged molecular inversion probes. en. *Nat. Biotechnol.* **21**, 673–678 (June 2003).
 104. Gunderson, K. L., Kruglyak, S., Graige, M. S., Garcia, F., Kermani, B. G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., Doucet, D., Milewski, M., Yang, R., Siegmund, C., Haas, J., Zhou, L., Oliphant, A., Fan, J.-B., Barnard,

- S. & Chee, M. S. Decoding randomly ordered DNA arrays. en. *Genome Res.* **14**, 870–877 (May 2004).
105. Jian-Bing Fan, K. Z. 20210087625 (2021).
 106. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. en. *Nat. Genet.* **37**, 549–554 (May 2005).
 107. Fan, J.-b., Gunderson, K. L., Bibikova, M., Yeakley, J. M., Chen, J., Wickham Garcia, E., Lebruska, L. L., Laurent, M., Shen, R. & Barker, D. in *Methods in Enzymology* 57–73 (Academic Press, Jan. 2006).
 108. Sun, Y.-C., Chen, X., Fischer, S., Lu, S., Zhan, H., Gillis, J. & Zador, A. M. Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. en. *Nat. Neurosci.* **24**, 873–885 (June 2021).
 109. Alon, S., Goodwin, D. R., Sinha, A., Wassie, A. T., Chen, F., Daugharthy, E. R., Bando, Y., Kajita, A., Xue, A. G., Marrett, K., Prior, R., Cui, Y., Payne, A. C., Yao, C.-C., Suk, H.-J., Wang, R., Yu, C.-C. J., Tillberg, P., Reginato, P., Pak, N., Liu, S., Punthambaker, S., Iyer, E. P. R., Kohman, R. E., Miller, J. A., Lein, E. S., Lako, A., Cullen, N., Rodig, S., Helvie, K., Abravanel, D. L., Wagle, N., Johnson, B. E., Klughammer, J., Slyper, M., Waldman, J., Jané-Valbuena, J., Rozenblatt-Rosen, O., Regev, A., IMAXT Consortium, Church, G. M., Marblestone, A. H. & Boyden, E. S. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. en. *Science* **371** (Jan. 2021).
 110. Diep, D., Plongthongkum, N., Gore, A., Fung, H.-L., Shoemaker, R. & Zhang, K. Library-free methylation sequencing with bisulfite padlock probes. en. *Nat. Methods* **9**, 270–272 (Feb. 2012).
 111. Chen, F., Tillberg, P. W. & Boyden, E. S. Optical imaging. Expansion microscopy. en. *Science* **347**, 543–548 (Jan. 2015).

112. Chen, F., Wassie, A. T., Cote, A. J., Sinha, A., Alon, S., Asano, S., Daugharthy, E. R., Chang, J.-B., Marblestone, A., Church, G. M., Raj, A. & Boyden, E. S. Nanoscale imaging of RNA with expansion microscopy. en. *Nat. Methods* **13**, 679–684 (Aug. 2016).
113. Bakken, T. E., Jorstad, N. L., Hu, Q., Lake, B. B., Tian, W., Kalmbach, B. E., Crow, M., Hodge, R. D., Krienen, F. M., Sorensen, S. A., Eggermont, J., Yao, Z., Aeversmann, B. D., Aldridge, A. I., Bartlett, A., Bertagnolli, D., Casper, T., Castanon, R. G., Crichton, K., Daigle, T. L., Dalley, R., Dee, N., Dembrow, N., Diep, D., Ding, S.-L., Dong, W., Fang, R., Fischer, S., Goldman, M., Goldy, J., Graybuck, L. T., Herb, B. R., Hou, X., Kancherla, J., Kroll, M., Lathia, K., van Lew, B., Li, Y. E., Liu, C. S., Liu, H., Lucero, J. D., Mahurkar, A., McMillen, D., Miller, J. A., Moussa, M., Nery, J. R., Nicovich, P. R., Niu, S.-Y., Orvis, J., Osteen, J. K., Owen, S., Palmer, C. R., Pham, T., Plongthongkum, N., Poirion, O., Reed, N. M., Rimorin, C., Rivkin, A., Romanow, W. J., Sedeño-Cortés, A. E., Siletti, K., Somasundaram, S., Sulc, J., Tieu, M., Torkelson, A., Tung, H., Wang, X., Xie, F., Yanny, A. M., Zhang, R., Ament, S. A., Behrens, M. M., Bravo, H. C., Chun, J., Dobin, A., Gillis, J., Hertzano, R., Hof, P. R., Höllt, T., Horwitz, G. D., Keene, C. D., Kharchenko, P. V., Ko, A. L., Lelieveldt, B. P., Luo, C., Mukamel, E. A., Pinto-Duarte, A., Preissl, S., Regev, A., Ren, B., Scheuermann, R. H., Smith, K., Spain, W. J., White, O. R., Koch, C., Hawrylycz, M., Tasic, B., Macosko, E. Z., McCarroll, S. A., Ting, J. T., Zeng, H., Zhang, K., Feng, G., Ecker, J. R., Linnarsson, S. & Lein, E. S. Comparative cellular analysis of motor cortex in human, marmoset and mouse. en. *Nature* **598**, 111–119 (Oct. 2021).
114. Yao, Z., Liu, H., Xie, F., Fischer, S., Adkins, R. S., Aldridge, A. I., Ament, S. A., Bartlett, A., Behrens, M. M., Van den Berge, K., Bertagnolli, D., de Bézieux, H. R., Biancalani, T., Booeshaghi, A. S., Bravo, H. C., Casper, T., Colantuoni, C., Crabtree, J., Creasy, H., Crichton, K., Crow, M., Dee, N., Dougherty, E. L., Doyle, W. I., Dudoit, S., Fang, R., Felix, V., Fong, O., Giglio, M., Goldy, J.,

- Hawrylycz, M., Herb, B. R., Hertzano, R., Hou, X., Hu, Q., Kancherla, J., Kroll, M., Lathia, K., Li, Y. E., Lucero, J. D., Luo, C., Mahurkar, A., McMillen, D., Nadaf, N. M., Nery, J. R., Nguyen, T. N., Niu, S.-Y., Ntranos, V., Orvis, J., Osteen, J. K., Pham, T., Pinto-Duarte, A., Poirion, O., Preissl, S., Purdom, E., Rimorin, C., Risso, D., Rivkin, A. C., Smith, K., Street, K., Sulc, J., Svensson, V., Tieu, M., Torkelson, A., Tung, H., Vaishnav, E. D., Vanderburg, C. R., van Velthoven, C., Wang, X., White, O. R., Huang, Z. J., Kharchenko, P. V., Pachter, L., Ngai, J., Regev, A., Tasic, B., Welch, J. D., Gillis, J., Macosko, E. Z., Ren, B., Ecker, J. R., Zeng, H. & Mukamel, E. A. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. en. *Nature* **598**, 103–110 (Oct. 2021).
115. Zhang, M., Eichhorn, S. W., Zingg, B., Yao, Z., Cotter, K., Zeng, H., Dong, H. & Zhuang, X. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. en. *Nature* **598**, 137–143 (Oct. 2021).
116. Gray, D. A. & Woulfe, J. Lipofuscin and aging: a matter of toxic waste. en. *Sci. Aging Knowledge Environ.* **2005**, re1 (Feb. 2005).
117. White, K., Yang, P., Li, L., Farshori, A., Medina, A. E. & Zielke, H. R. Effect of Postmortem Interval and Years in Storage on RNA Quality of Tissue at a Repository of the NIH NeuroBioBank. en. *Biopreserv. Biobank.* **16**, 148–157 (Apr. 2018).
118. Müller, C., Bauer, N. M., Schäfer, I. & White, R. Making myelin basic protein -from mRNA transport to localized translation. en. *Front. Cell. Neurosci.* **7**, 169 (Sept. 2013).
119. Costa, M. R. & Müller, U. Specification of excitatory neurons in the developing cerebral cortex: progenitor diversity and environmental influences. en. *Front. Cell. Neurosci.* **8**, 449 (2014).
120. Nieto, M., Monuki, E. S., Tang, H., Imitola, J., Haubst, N., Khoury, S. J., Cunningham, J., Gotz, M. & Walsh, C. A. Expression of Cux-1 and Cux-2 in

- the subventricular zone and upper layers II-IV of the cerebral cortex. en. *J. Comp. Neurol.* **479**, 168–180 (Nov. 2004).
121. Schaeren-Wierners, N., André, E., Kapfhammer, J. P. & Becker-André, M. The Expression pattern of the orphan nuclear receptor ROR β in the developing and adult rat nervous system suggests a role in the processing of sensory information and in circadian rhythm. en. *Eur. J. Neurosci.* **9**, 2687–2701 (Dec. 1997).
 122. Arlotta, P., Molyneaux, B. J., Chen, J., Inoue, J., Kominami, R. & Macklis, J. D. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. en. *Neuron* **45**, 207–221 (Jan. 2005).
 123. Reiner, A., Yang, M., Cagle, M. C. & Honig, M. G. Localization of cerebellin-2 in late embryonic chicken brain: implications for a role in synapse formation and for brain evolution. en. *J. Comp. Neurol.* **519**, 2225–2251 (Aug. 2011).
 124. Fang, R., Xia, C., Close, J. L., Zhang, M., He, J., Huang, Z., Halpern, A. R., Long, B., Miller, J. A., Lein, E. S. & Zhuang, X. Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. en. *Science* **377**, 56–62 (July 2022).
 125. Borm, L. E., Mossi Albiach, A., Mannens, C. C. A., Janusauskas, J., Özgün, C., Fernández-García, D., Hodge, R., Castillo, F., Hedin, C. R. H., Villablanca, E. J., Uhlén, P., Lein, E. S., Codeluppi, S. & Linnarsson, S. Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. en. *Nat. Biotechnol.* **41**, 222–231 (Feb. 2023).
 126. Jorstad, N. L., Close, J., Johansen, N., Yanny, A. M., Barkan, E. R., Travaglini, K. J., Bertagnolli, D., Campos, J., Casper, T., Crichton, K., Dee, N., Ding, S.-L., Gelfand, E., Goldy, J., Hirschstein, D., Kroll, M., Kunst, M., Lathia, K., Long, B., Martin, N., McMillen, D., Pham, T., Rimorin, C., Ruiz, A., Shapovalova, N., Shehata, S., Siletti, K., Somasundaram, S., Sulc, J., Tieu, M., Torkelson, A., Tung, H., Ward, K., Callaway, E. M., Hof, P. R., Dirk Keene, C., Levi, B. P., Linnarsson, S., Mitra, P. P., Smith, K., Hodge, R. D., Bakken, T. E. & Lein,

- E. S. *Transcriptomic cytoarchitecture reveals principles of human neocortex organization* en. Nov. 2022.
127. Jorstad, N. L., Song, J. H. T., Exposito-Alonso, D., Suresh, H., Castro, N., Krienen, F. M., Yanny, A. M., Close, J., Gelfand, E., Travaglini, K. J., Basu, S., Beaudin, M., Bertagnolli, D., Crow, M., Ding, S.-L., Eggermont, J., Glandon, A., Goldy, J., Kroes, T., Long, B., McMillen, D., Pham, T., Rimorin, C., Siletti, K., Somasundaram, S., Tieu, M., Torkelson, A., Ward, K., Feng, G., Hopkins, W. D., Höllt, T., Dirk Keene, C., Linnarsson, S., McCarroll, S. A., Lelieveldt, B. P., Sherwood, C. C., Smith, K., Walsh, C. A., Dobin, A., Gillis, J., Lein, E. S., Hodge, R. D. & Bakken, T. E. *Comparative transcriptomics reveals human-specific cortical features* en. Sept. 2022.
 128. Schmitz, M. T., Sandoval, K., Chen, C. P., Mostajo-Radji, M. A., Seeley, W. W., Nowakowski, T. J., Ye, C. J., Paredes, M. F. & Pollen, A. A. The development and evolution of inhibitory neurons in primate cerebrum. en. *Nature* **603**, 871–877 (Mar. 2022).
 129. Bandler, R. C. & Mayer, C. Deciphering inhibitory neuron development: The paths to diversity. en. *Curr. Opin. Neurobiol.* **79**, 102691 (Apr. 2023).
 130. Kim, Y., Yang, G. R., Pradhan, K., Venkataraju, K. U., Bota, M., Garcia Del Molino, L. C., Fitzgerald, G., Ram, K., He, M., Levine, J. M., Mitra, P., Huang, Z. J., Wang, X.-J. & Osten, P. Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. en. *Cell* **171**, 456–469.e22 (Oct. 2017).
 131. Chun, J. J. & Shatz, C. J. The earliest-generated neurons of the cat cerebral cortex: characterization by MAP2 and neurotransmitter immunohistochemistry during fetal life. en. *J. Neurosci.* **9**, 1648–1667 (May 1989).
 132. Chun, J. J. & Shatz, C. J. Interstitial cells of the adult neocortical white matter are the remnant of the early generated subplate neuron population. en. *J. Comp. Neurol.* **282**, 555–569 (Apr. 1989).

133. Tan, S.-S., Kalloniatis, M., Truong, H.-T., Binder, M. D., Cate, H. S., Kilpatrick, T. J. & Hammond, V. E. Oligodendrocyte positioning in cerebral cortex is independent of projection neuron layering. en. *Glia* **57**, 1024–1030 (July 2009).
134. Ohtomo, R., Iwata, A. & Arai, K. Molecular Mechanisms of Oligodendrocyte Regeneration in White Matter-Related Diseases. en. *Int. J. Mol. Sci.* **19** (June 2018).
135. Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., Hawrylycz, M., Koch, C. & Zeng, H. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. en. *Nat. Neurosci.* **19**, 335–346 (Feb. 2016).
136. Bulfone, A., Caccioppoli, C., Pardini, C., Faedo, A., Martinez, S. & Banfi, S. *Pcp4l1*, a novel gene encoding a *Pcp4*-like polypeptide, is expressed in specific domains of the developing brain. en. *Gene Expr. Patterns* **4**, 297–301 (May 2004).
137. Axelrod, S., Cai, M., Carr, A., Freeman, J., Ganguli, D., Kiggins, J., Long, B., Tung, T. & Yamauchi, K. starfish: scalable pipelines for image-based transcriptomics. *J. Open Source Softw.* **6**, 2440 (May 2021).
138. Andersson, A., Diego, F., Hamprecht, F. A. & Wählby, C. *ISTDECO: In Situ Transcriptomics Decoding by Deconvolution* en. Mar. 2021.
139. Chen, S., Loper, J., Chen, X., Vaughan, A., Zador, A. M. & Paninski, L. BAR-code DEmixing through Non-negative Spatial Regression (BarDensr). en. *PLoS Comput. Biol.* **17**, e1008256 (Mar. 2021).
140. Sainio, A. & Järveläinen, H. Extracellular matrix-cell interactions: Focus on therapeutic applications. en. *Cell. Signal.* **66**, 109487 (Feb. 2020).

141. Wang, Y., Eddison, M., Fleishman, G., Weigert, M., Xu, S., Wang, T., Rokicki, K., Goina, C., Henry, F. E., Lemire, A. L., Schmidt, U., Yang, H., Svoboda, K., Myers, E. W., Saalfeld, S., Korff, W., Sternson, S. M. & Tillberg, P. W. EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization. en. *Cell* **184**, 6361–6377.e24 (Dec. 2021).
142. Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Jacob Filho, W., Lent, R. & Herculano-Houzel, S. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. en. *J. Comp. Neurol.* **513**, 532–541 (Apr. 2009).
143. Von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. en. *J. Comp. Neurol.* **524**, 3865–3895 (Dec. 2016).
144. Sunkin, S. M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T. L., Thompson, C. L., Hawrylycz, M. & Dang, C. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. en. *Nucleic Acids Res.* **41**, D996–D1008 (Jan. 2013).
145. Langseth, C. M., Gyllborg, D., Miller, J. A., Close, J. L., Long, B., Lein, E. S., Hilscher, M. M. & Nilsson, M. Comprehensive in situ mapping of human cortical transcriptomic cell types. en. *Commun Biol* **4**, 998 (Aug. 2021).
146. Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., Tokcan, N., Vanderburg, C. R., Segerstolpe, Å., Zhang, M., Avraham-Davidi, I., Vickovic, S., Nitzan, M., Ma, S., Subramanian, A., Lipinski, M., Buenrostro, J., Brown, N. B., Fanelli, D., Zhuang, X., Macosko, E. Z. & Regev, A. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. en. *Nat. Methods* **18**, 1352–1362 (Nov. 2021).
147. Ding, S.-L., Royall, J. J., Sunkin, S. M., Ng, L., Facer, B. A. C., Lesnar, P., Guillozet-Bongaarts, A., McMurray, B., Szafer, A., Dolbeare, T. A., Stevens,

- A., Tirrell, L., Benner, T., Caldejon, S., Dalley, R. A., Dee, N., Lau, C., Nyhus, J., Reding, M., Riley, Z. L., Sandman, D., Shen, E., van der Kouwe, A., Varjabedian, A., Write, M., Zollei, L., Dang, C., Knowles, J. A., Koch, C., Phillips, J. W., Sestan, N., Wohnoutka, P., Zielke, H. R., Hohmann, J. G., Jones, A. R., Bernard, A., Hawrylycz, M. J., Hof, P. R., Fischl, B. & Lein, E. S. Comprehensive cellular-resolution atlas of the adult human brain. *J. Comp. Neurol.* **524**, 3127–3481 (Nov. 2016).
148. Zhang, Y., Petukhov, V., Biederstedt, E., Que, R., Zhang, K. & Kharchenko, P. V. Gene panel selection for targeted spatial transcriptomics. en. *bioRxiv* (Mar. 2023).
 149. Garcia-Cabezas, M. Á. & Barbas, H. Area 4 has layer IV in adult primates. en. *Eur. J. Neurosci.* **39**, 1824–1834 (June 2014).
 150. Yamawaki, N., Borges, K., Suter, B. A., Harris, K. D. & Shepherd, G. M. G. A genuine layer 4 in motor cortex with prototypical synaptic circuit connectivity. en. *Elife* **3**, e05422 (Dec. 2014).
 151. Barbas, H. & Garcia-Cabezas, M. Á. Motor cortex layer 4: less is more. en. *Trends Neurosci.* **38**, 259–261 (May 2015).
 152. Mo, Z. & Zecevic, N. Is Pax6 critical for neurogenesis in the human fetal brain? en. *Cereb. Cortex* **18**, 1455–1465 (June 2008).
 153. Osumi, N., Shinohara, H., Numayama-Tsuruta, K. & Maekawa, M. Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. en. *Stem Cells* **26**, 1663–1672 (July 2008).
 154. Georgala, P. A., Manuel, M. & Price, D. J. The generation of superficial cortical layers is regulated by levels of the transcription factor Pax6. en. *Cereb. Cortex* **21**, 81–94 (Jan. 2011).
 155. Simons, M. & Nave, K.-A. Oligodendrocytes: Myelination and Axonal Support. en. *Cold Spring Harb. Perspect. Biol.* **8**, a020479 (June 2015).

156. Volterra, A. & Meldolesi, J. Astrocytes, from brain glue to communication elements: the revolution continues. en. *Nat. Rev. Neurosci.* **6**, 626–640 (Aug. 2005).
157. Siracusa, R., Fusco, R. & Cuzzocrea, S. Astrocytes: Role and Functions in Brain Pathologies. en. *Front. Pharmacol.* **10**, 1114 (Sept. 2019).
158. Kalmbach, B. E., Hodge, R. D., Jorstad, N. L., Owen, S., de Frates, R., Yanny, A. M., Dalley, R., Mallory, M., Graybuck, L. T., Radaelli, C., Keene, C. D., Gwinn, R. P., Silbergeld, D. L., Cobbs, C., Ojemann, J. G., Ko, A. L., Patel, A. P., Ellenbogen, R. G., Bakken, T. E., Daigle, T. L., Dee, N., Lee, B. R., McGraw, M., Nicovich, P. R., Smith, K., Sorensen, S. A., Tasic, B., Zeng, H., Koch, C., Lein, E. S. & Ting, J. T. Signature morpho-electric, transcriptomic, and dendritic properties of human layer 5 neocortical pyramidal neurons. en. *Neuron* **109**, 2914–2927.e5 (Sept. 2021).
159. Fries, W., Keizer, K. & Kuypers, H. G. Large layer VI cells in macaque striate cortex (Meynert cells) project to both superior colliculus and prestriate visual area V5. en. *Exp. Brain Res.* **58**, 613–616 (1985).
160. Nhan, H. L. & Callaway, E. M. Morphology of superior colliculus- and middle temporal area-projecting neurons in primate primary visual cortex. en. *J. Comp. Neurol.* **520**, 52–80 (Jan. 2012).
161. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R. & Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. en. *Nat. Methods* **16**, 1289–1296 (Dec. 2019).
162. Sohal, V. S. & Rubenstein, J. L. R. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. en. *Mol. Psychiatry* **24**, 1248–1257 (Sept. 2019).

163. Ueda, H. R., Ertürk, A., Chung, K., Gradinaru, V., Chédotal, A., Tomancak, P. & Keller, P. J. Tissue clearing and its applications in neuroscience. en. *Nat. Rev. Neurosci.* **21**, 61–79 (Feb. 2020).
164. Ueda, H. R., Dodt, H.-U., Osten, P., Economo, M. N., Chandrashekar, J. & Keller, P. J. Whole-Brain Profiling of Cells and Circuits in Mammals by Tissue Clearing and Light-Sheet Microscopy. en. *Neuron* **106**, 369–387 (May 2020).
165. Sen, S. Q., Chanchani, S., Southall, T. D. & Doe, C. Q. Neuroblast-specific open chromatin allows the temporal transcription factor, Hunchback, to bind neuroblast-specific loci. en. *Elife* **8** (Jan. 2019).
166. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997 [q-bio.GN] (Mar. 2013).
167. Lowekamp, B., gabehart, Blezek, D., Ibanez, L., McCormick, M., Chen, D., Mueller, D., Cole, D., Johnson, H., Marstal, K., Beare, R., Gelas, A., Williams, K., Doria, D. & King, B. *SimpleElastix: SimpleElastix v0.9.0* June 2015.
168. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. Fiji: an open-source platform for biological-image analysis. en. *Nat. Methods* **9**, 676–682 (June 2012).
169. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. en. *Bioinformatics* **25**, 1463–1465 (June 2009).
170. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. en. *Nat. Methods* **18**, 100–106 (Jan. 2021).
171. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. en. *Nat. Methods* **19**, 1634–1641 (Dec. 2022).
172. Efron, B. & Hastie, T. *Computer Age Statistical Inference, Student Edition: Algorithms, Evidence, and Data Science* en (Cambridge University Press, June 2021).

173. Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T. & scikit-image contributors. scikit-image: image processing in Python. en. *PeerJ* **2**, e453 (June 2014).
174. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. en. *Genome Biol.* **19**, 15 (Feb. 2018).
175. Virshup, I., Bredikhin, D., Heumos, L., Palla, G., Sturm, G., Gayoso, A., Kats, I., Koutrouli, M., Scverse Community, Berger, B., Pe’er, D., Regev, A., Teichmann, S. A., Finotello, F., Wolf, F. A., Yosef, N., Stegle, O. & Theis, F. J. The scverse project provides a computational ecosystem for single-cell omics data analysis. en. *Nat. Biotechnol.* **41**, 604–606 (May 2023).
176. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Alexander Wolf, F. *anndata: Annotated data* en. Dec. 2021.
177. Lause, J., Berens, P. & Kobak, D. Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. en. *Genome Biol.* **22**, 258 (Sept. 2021).
178. Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. en. *Genome Biol.* **23**, 27 (Jan. 2022).
179. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. en. *Sci. Rep.* **9**, 5233 (Mar. 2019).