Optics-free reconstruction of 2D images via DNA barcode proximity graphs

Hanna Liao[1,2,*], Sanjay Kottapalli[1,*], Yuqi Huang[3], Matthew Chaw[1,3], Jase Gehring[1], Olivia Waltner[1], Melissa Phung-Rojas[3], Riza M. Daza[1,6], Frederick A. Matsen IV[3,4,5,7], Cole Trapnell[1,6,8], Jay Shendure[1,6,7,8,#], Sanjay Srivatsan[1,3,8,#]

[1] Department of Genome Sciences, University of Washington, Seattle, WA, USA
[2] Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA
[3] Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[4] Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA, USA
[5] Department of Statistics, University of Washington, Seattle, WA, USA
[6] Seattle Hub for Synthetic Biology, Seattle, WA, USA
[7] Howard Hughes Medical Institute, Seattle, WA, USA
[8] Brotman Baty Institute for Precision Medicine, Seattle, WA, USA

* These authors contributed equally to this work.
[#] To whom correspondence should be addressed. Email: shendure@uw.edu, sanjay@fredhutch.org

Abstract

Spatial genomic technologies include imaging- and sequencing-based methods (*1–3*). An emerging subcategory of sequencing-based methods relies on a surface coated with coordinate-associated DNA barcodes, which are leveraged to tag endogenous nucleic acids or cells in an overlaid tissue section (*4–7*). However, the physical registration of DNA barcodes to spatial coordinates is challenging, necessitating either high density printing of coordinate-specific oligonucleotides or *in situ* sequencing/probing of randomly deposited, oligonucleotide-bearing beads. As a consequence, the surface areas available to sequencing-based spatial genomic methods are constrained by the time, labor, cost, and instrumentation required to either print, synthesize or decode a coordinate-tagged surface. To address this challenge, we developed SCOPE (Spatial reConstruction via Oligonucleotide Proximity Encoding), an optics-free, DNA microscopy (*8*) inspired method. With SCOPE, the relative positions of randomly deposited beads on a 2D surface are inferred from the *ex situ* sequencing of chimeric molecules formed from diffusing “sender” and tethered “receiver” oligonucleotides. As a first proof-of-concept, we apply SCOPE to reconstruct an asymmetric “swoosh” shape resembling the Nike logo (16.75 x 9.25 mm). Next, we use a microarray printer to encode a “color” version of the Snellen eye chart for visual acuity (17.18 x 40.97 mm), and apply SCOPE to achieve optics-free reconstruction of individual letters. Although these are early demonstrations of the concept and much work remains to be done, we envision that the optics-free, sequencing-based quantitation of the molecular proximities of DNA barcodes will enable spatial genomics in constant experimental time, across fields of view and at resolutions that are determined by sequencing depth, bead size, and diffusion kinetics, rather than the limitations of optical instruments or microarray printers.

Introduction

In the natural and engineered worlds, conventional paradigms for determining spatial relationships include optics (*e.g.* microscopy), acoustics (*e.g.* echolocation) and haptics (*e.g.* touch). However, particularly at nanoscopic and microscopic scales, one can also imagine how complex polymers ─ DNA in particular ─ might be leveraged to encode and decode spatial information. For such a goal, DNA has a number of attractive properties, including its high information density, simple design principles, biochemical stability; the breadth of primitive operations that can be accessed enzymatically or directly encoded (*9*); and our contemporary access to low-cost technologies for synthesizing and sequencing nucleic acids of arbitrary sequence composition. Many of these same properties underlie DNA's promise as a digital data storage medium, whether *in vitro* (*e.g.* using A's, G's, C's and T's to encode written symbols or visual pixels (*10*, *11*)) or *in vivo* (*e.g.* using prime editing-mediated short insertions to encode ordered molecular signaling or cell lineage histories (*12*, *13*)).

DNA-based molecular barcoding (aka tagging or multiplexing) (*14*) is at the core of many modern genomic methods, including unique molecular identifiers (UMI) [in which RNAs are tagged], massively parallel reporter assays (MPRAs) [in which regulatory elements are tagged], pooled genetic screens [in which perturbations, variants or lineages are tagged], single cell combinatorial indexing [in which cells are tagged], proximity ligation assays [in which proteins or nucleic acids are tagged] and spatial genomics [in which hybridization probes or spatial coordinates are tagged]. In these methods, the barcode sequences are mostly arbitrary, while for some further applications (*e.g.* DNA origami (*15*), DNA computing (*16*)), barcode interactions encode molecular logic.

Over the past decade, spatial genomic technologies have rapidly proliferated (*1–3*). The underlying methods can be dichotomized into those that rely on *in situ* imaging vs. *ex situ* sequencing for primary data collection, while *ex situ* sequencing methods can be further dichotomized into those that rely on microdissection vs. DNA-based molecular barcodes to identify the spatial coordinates associated with each sequencing read. The main advantages of *ex situ* sequencing methods include that they do not require *a priori* specification of molecular species of interest and are less dependent on the cost and throughput limitations of optical instrumentation, while a further advantage of the spatial barcoding subset of *ex situ* sequencing methods is that substantially greater resolution can be achieved than with microdissection.

However, a major limitation of spatial barcoding methods is that they require the deposition and mapping of specific barcoded oligonucleotides to specific locations on a 2D surface. The mapping between specific barcodes and specific physical coordinates is analogous to the mapping between the elements of a CCD array and individual pixels in a digital camera. To fabricate and map arrays of DNA barcodes, oligonucleotides can be: **1)** pre-synthesized, arrayed, and deposited to specific locations, *e.g.* by a microarray printer (*6*, *7*); **2)** synthesized to specific locations *in situ (17)*; or **3)** randomly distributed and decoded via *in situ* hybridization or sequencing (*5*, *18*, *19*). However, these approaches are all time-, cost-, labor- and/or instrumentation-intensive, which limits the physical dimensions of the arrays that are typically produced and used for these assays.

DNA microscopy is a highly innovative method pioneered by Weinstein *et al. (8)*. In brief, endogenous transcripts in fixed cells or tissues are tagged with UMIs, amplified, concatenated and sequenced. As the probability of concatenation is dependent on molecular proximity, the physical relationships among the original transcripts can be computationally inferred with cellular resolution, *i.e.* a "chemically encoded microscopy system" (*8*). Several groups have developed novel computational algorithms for DNA microscopy data (*20–22*), and it has recently been adapted to spatial proteomics (*23*) as well as to achieve optics-free "imaging" of

developing zebrafish embryos (*24*). However, as a spatial genomics technology, DNA microscopy remains limited in terms of resolution and sensitivity.

Here, inspired by DNA microscopy, we developed SCOPE (Spatial reConstruction via Oligonucleotide Proximity Encoding), an optics-free method in which a 2D array of randomly deposited beads, each bearing a unique DNA barcode, essentially self-register their relative positions via proximity-dependent hybridization and overlap extension of released "sender" and tethered "receiver" oligonucleotides. Massively parallel sequencing of sender-receiver chimeras results in a matrix of pairwise counts. Particularly for local communities, applying dimensionality reduction algorithms widely used for scRNA-seq analysis, t-Distributed Stochastic Neighbor Embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP), to such data results in reasonable reconstructions of the relative positions of thousands of beads (*25*, *26*). As a first proof-of-concept, we apply SCOPE to reconstruct an asymmetric "swoosh" resembling the Nike logo. As a second proof-of-concept, we use oligonucleotides and a microarray printer to encode a multi-color image of the Snellen eye chart for visual acuity, which we then attempt to reconstruct in an optics-free manner with SCOPE.

Results

Overview of SCOPE

We set out to develop a method for generating arbitrarily large surfaces of DNA barcodes whose relative spatial positions could be determined without relying on optical or microjet instrumentation. The strategy that we devised is based on a population of beads bearing unique DNA barcodes, *i.e.* wherein any given bead bears many copies of a single DNA barcode, while different beads bear other DNA barcodes. However, the oligonucleotides borne by any given bead are still heterogeneous in a different sense, in that a subset are "messengers", designed to interact with messengers from nearby beads, while the remainder are "decoders", designed to capture molecules of interest, *e.g.* polyA-tailed mRNA or ssDNA species (**Fig. 1A**). Moreover, within each of these categories, further subsets are designed for programmed release ("senders") or to remain tethered ("receivers") (**Fig. 1A**). After generating a 2D surface of densely packed, DNA barcode-bearing beads, sender-messengers are released to diffuse, hybridize and chimerize with receiver-messengers that remain tethered to nearby beads in the x-y plane (**Fig. 1B**). At the same time, decoder molecules can either diffuse out to find (sender-decoders) or stay tethered to receive (receiver-decoders) molecules of interest that are proximal to a given bead in the z plane, *e.g.* in an overlaid tissue section or printed oligo image (**Fig. 1B**). Massively parallel sequencing of sender-receiver messenger chimeras is expected to result in a matrix of barcode-barcode interactions that is informative with respect to the proximity relationships of beads (**Fig. 1C**), while massively parallel sequencing of chimeras involving decoders is expected to map molecules-of-interest to the proximity of specific beads (**Fig. 1C**).

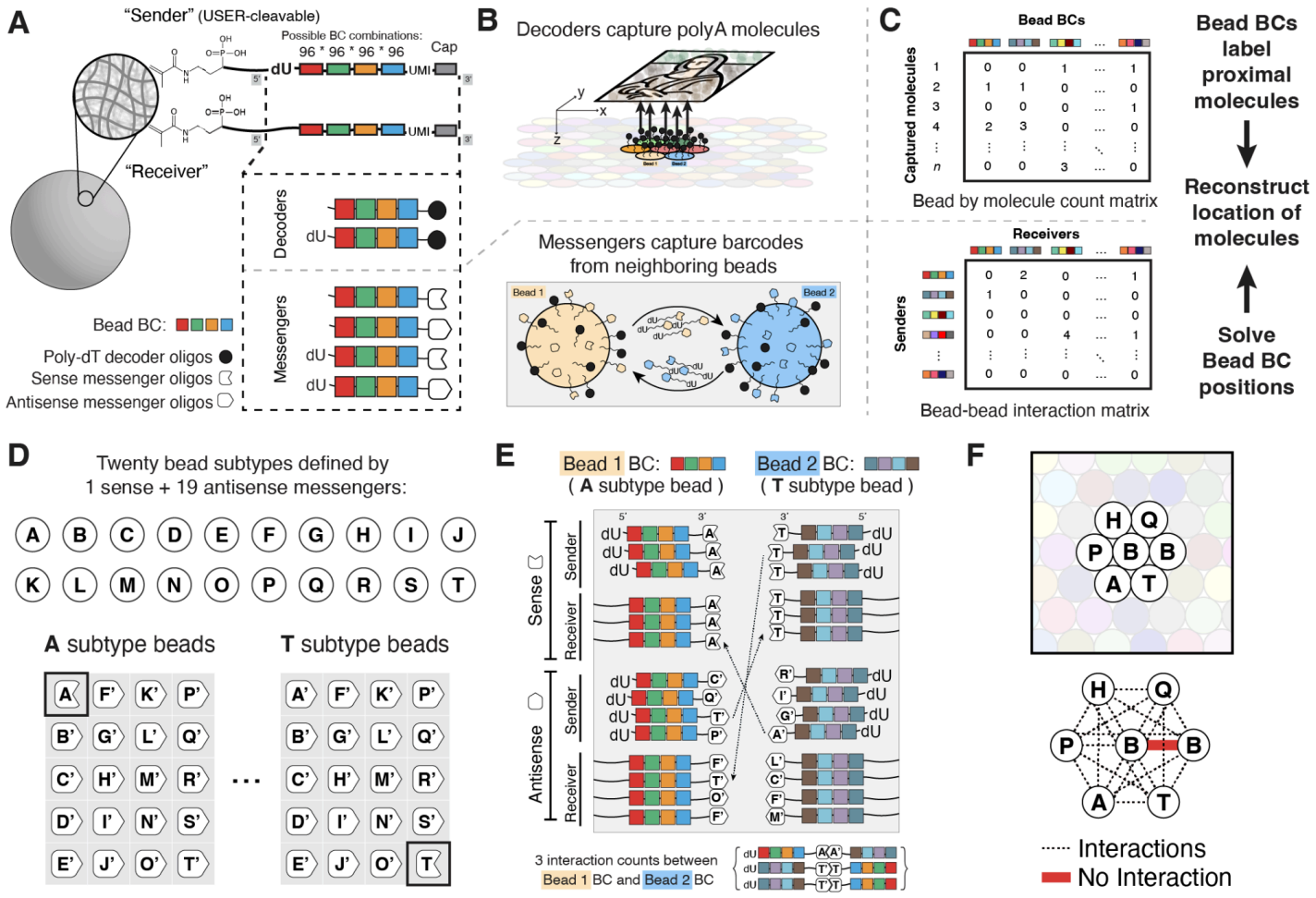


Figure 1. Schematic of Spatial reConstruction via Oligonucleotide Proximity Encoding (SCOPE). (A) Hydrogel beads contain 5'-Acrydite-tethered oligonucleotides that are either USER enzyme cleavable (senders) or non-cleavable (receivers) based on the presence/absence of dU at the stem. Both species contain the same DNA barcode – a combination of four positions with 96 possible sequences at each position – depicted by the four colored boxes. Each DNA barcode is functionalized with a capping sequence of either poly-dT for capturing polyA-tailed molecules (decoders) or a subtype-defining sequence in one orientation or the other for participating in proximity reactions between beads (messengers). **(B)** Illustration of the SCOPE reaction to achieve both spatial reconstruction and capture of proximal molecules-of-interest. *Top:* Beads arrayed in a 2D plane are proximate to polyA-tailed molecules of interest in an adjacent plane. Decoder molecules are designed to form chimeras and thereby label proximal molecules-of-interest with a spatially informative DNA barcode. *Bottom*: Messenger molecules are designed to participate in diffusion-dependent crosstalk with the messengers of proximal beads in the 2D surface, resulting in sender-receiver messenger chimeras that are informative with respect to bead-bead proximity. **(C)** *Top:* From the spatial labeling of proximal polyA-tailed molecules via decoders, a count matrix of bead DNA barcodes-by-captured molecules will be generated. *Bottom:* From proximity reactions between neighboring beads via messengers, a count matrix of bead DNA barcodes-by-bead DNA barcodes will be generated. The former informs spatial mapping of molecules-of-interest, while the latter informs the inference of the relative positions of beads. **(D)** Each of twenty bead subtypes contains a distinct combination of sense and antisense messenger oligonucleotides. Of the twenty pairs of complementary messenger sequences, the barcodes on each bead are capped with 19 "antisense" sequences and one "sense" sequence. **(E)** Depiction of three proximity counts originating from the interaction of the two adjacent beads, where the matching sense/antisense pairs form chimeras. The DNA barcode from the sender-messenger is always on the 5' end of the chimeric sender-receiver molecule. **(F)** Only beads of different subtypes can contribute to the generation of sender-receiver messenger chimeras.

Generation of large numbers of combinatorially barcoded, SCOPE-enabled beads

We chose to use 20 μm hydrogel beads as a solid substrate for SCOPE's proof-of-concept, as accessible, low-cost methods exist for scaled production of such beads bearing unique DNA barcodes (*27*) and for generating densely packed, large 2D bead arrays. Acrylamide hydrogels have a high internal porosity which facilitates a higher density loading of oligonucleotides compared to non-porous 2-D supports like glass or polystyrene (*19*). We generated 20 μm polyacrylamide hydrogel beads containing 100 μM of 5'-Acrydite oligos using a flow-focusing microfluidics chip (**Fig. S1A**). The 5'-Acrydite oligos were synthesized either with or without a deoxyuracil (dU) at the stem. The presence of the dU in only some 5'-Acrydite oligos allows for USER enzyme-controlled cleavage of a subset of DNA barcodes (*i.e.* "senders"), while its absence ensures that the remaining DNA barcodes remain tethered (*i.e.* "receivers"), during the SCOPE reaction (**Fig. 1A**). The 5'-Acrydite oligos act as handles for assembling the rest of the DNA barcode, which is achieved via a split-pool approach developed by Delley & Abate (*27*). More specifically, four rounds of splitting beads to 96 wells, while ligating well-specific sub-barcodes to bead-tethered oligonucleotides, results in a combinatorial space of $(96)^4$ or 85 million possible barcodes. In practice, we found the Delley & Abate combinatorial barcoding procedure (*27*) to be straightforward to adopt, with high efficiency and relatively uniform usage of sub-barcodes at each round of assembly (**Fig. S1B,C**). The final barcode, built from four iterative rounds of sub-barcode ligation, was then capped with functional sequences to enable messenger and decoder functionalities (**Fig. 1A-B**). Based on amplification and sequencing of UMIs associated with barcodes enzymatically released from individual beads, we estimate that each bead bears on the order of 500,000 functionalized, barcoded oligonucleotides (mean: 491,811; IQR: 308,757 - 833,744; **Fig. S1D**). To generate a planar array that meets SCOPE's core requirement (*i.e.* specific DNA barcodes concentrated in specific locations), SCOPE beads were cast within an encasing hydrogel that polymerized directly onto a treated glass slide (**Fig. S1E,F**).

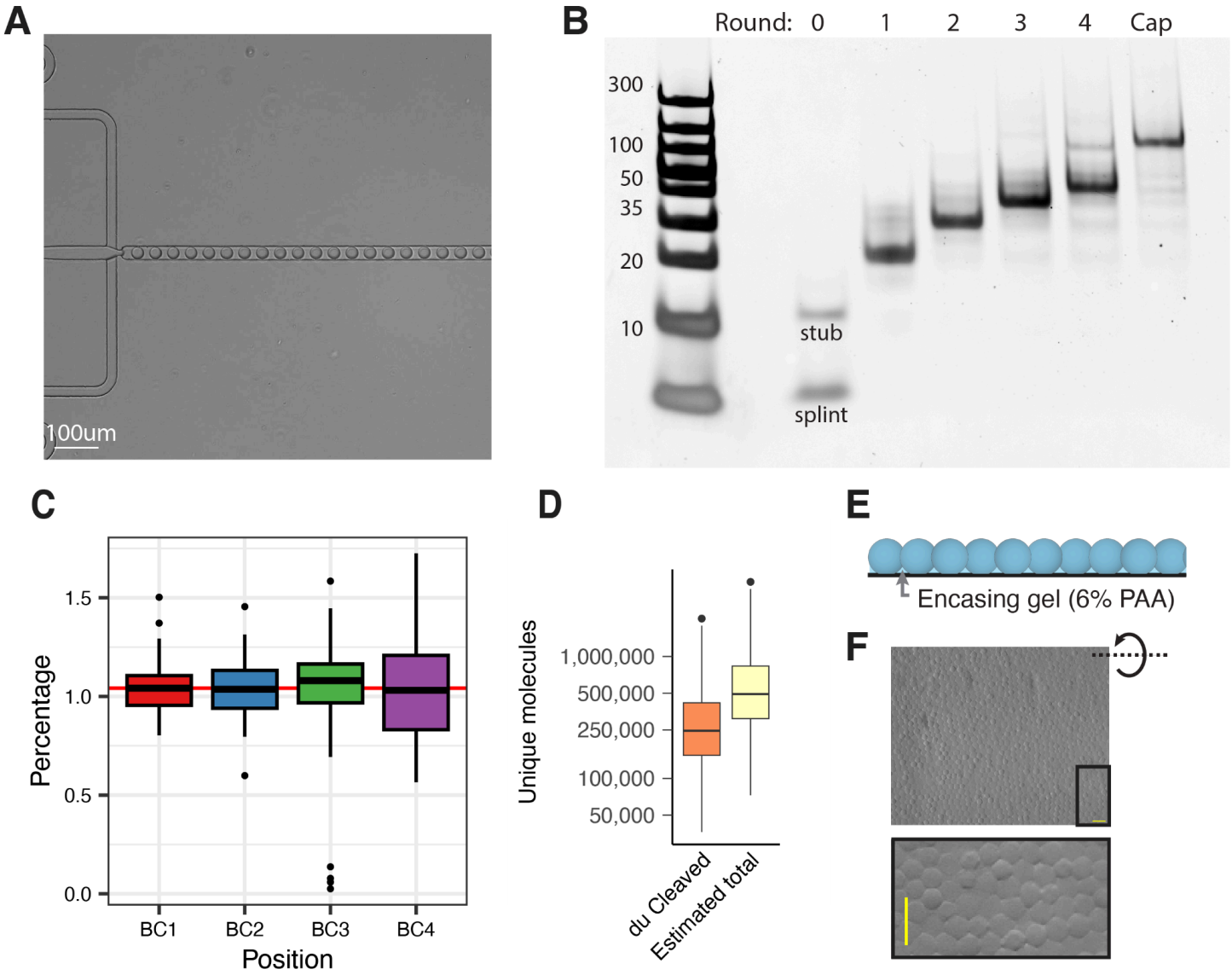
Enabling self-nonself interactions while minimizing self-self interactions among proximate beads

SCOPE requires that sender and receiver messengers have some region of complementarity (*i.e.* sense vs. antisense) to hybridize to one another. The scheme also bears some risk of "self-self" interactions, *i.e.* chimeras between senders and receiver messengers derived from the same bead, a challenge that also plagues some other proximity dependent methods, *e.g.* Hi-C. To minimize this risk, we designed 20 species to serve as the capping "functional sequence" of messengers during bead production (**Fig. 1A**; **Fig. S2A**). After synthesizing DNA barcodes but prior to capping, beads were split into 20 subsets and loaded with a mixture of messenger cap species, such that DNA barcodes in each subset or "bead subtype" were capped with the antisense strands of 19 of the messenger cap species, and the sense strand of the 1 remaining species (**Fig. 1D**). As a consequence, the sender or receiver messengers associated with a given bead subtype are capable of hybridizing with the receivers or senders, respectively, of all 19 other bead subtypes to give rise to informative chimeras (**Fig. 1E**). On the other hand, if messengers of two beads of the same subtype have the opportunity to interact, they will necessarily be non-complementary and unable to form chimeras (**Fig. 1E,F**). The choice of 20 bead subtypes was based on simulations indicating diminishing returns for additional bead subtypes in terms of minimizing the extent to which neighboring beads were of the same subtype (**Fig. S2B**).

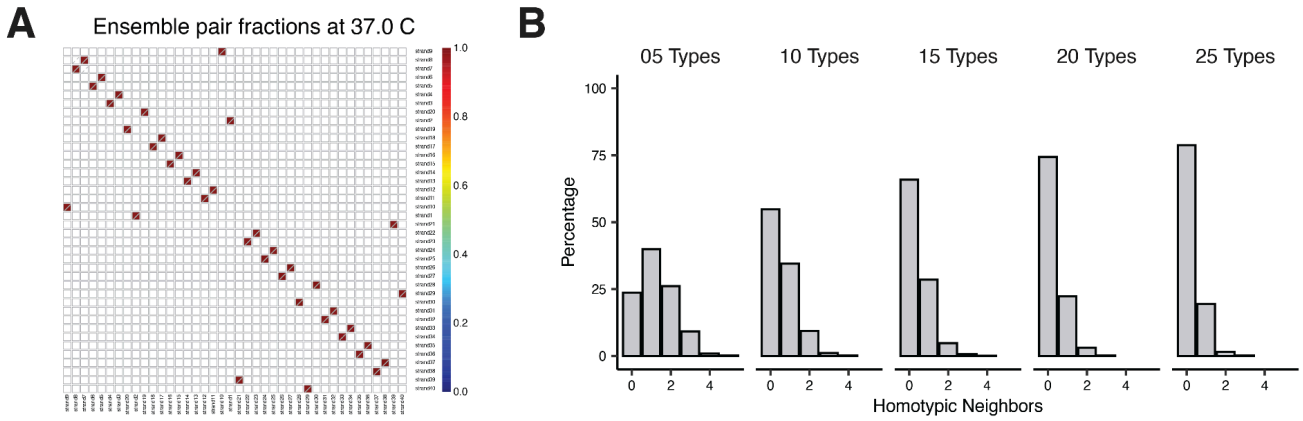
SCOPE reaction and massively parallel sequencing of sender-receiver chimeras

The SCOPE reaction involves subjecting the hydrogel-encased monolayer of polyacrylamide beads (**Fig. S1E,F**) to a temperature-controlled, two-phase reaction mediated by a single reaction volume containing two enzymes. In the first step (37°C; 15'), sender-messengers are released from beads by the USER enzyme,

which cleaves at the dU present at the stem of a subset of DNA barcodes (**Fig. 1A**). Sender-messengers diffuse away and hybridize to receiver-messengers, of the same bead type but opposite strand, tethered to adjacent beads (**Fig. 1B,E,F**). Because the complementary regions reside in the functional cap (**Fig. 1A**), senders and receivers are expected to hybridize at their 3' ends. In the second phase (55°C; 15'), reverse transcriptase, which mediates both RNA- and DNA-templated DNA polymerization, drives overlap extension and the generation of a chimeric molecule that includes both the sender-messenger and receiver-messenger derived DNA barcodes (or, alternatively, chimeras of decoders and proximate molecules-of-interest) (**Fig. 1B**). Chimeric molecules are then PCR amplified and subjected to massively parallel sequencing. The result is a sparse matrix of interaction counts between a large number of DNA barcode-defined beads. As the SCOPE reaction is dependent on a sender-messenger diffusing into the vicinity of a receiver-messenger, we expect the count frequencies between any given pair of beads to be a function of their physical distance. Importantly, the asymmetry of primer binding sites is such that for any given chimera, we are aware of which participating bead is the sender vs. the receiver. The resulting matrix of count frequencies is therefore asymmetric (**Fig. 1C**).



Supplementary Figure 1. Barcoded hydrogel bead generation and bead array fabrication. (A) Image of the flow focusing microfluidics generator while producing polyacrylamide hydrogel beads. **(B)** Oligos released by USER-mediated cleavage from the initial bead (0) vs. after each round of DNA synthesis by ligation of the barcode (1, 2, 3, 4) or functional cap (Cap), next to a DNA size ladder (bp), in a 10% polyacrylamide gel. In the lane corresponding to the initial bead, the stub is what is cleaved off the strand that was initially incorporated into the hydrogel bead, while the the splint is a primer that anneals to the bead stub to create a 4-bp overhang handle for the ligation of the first of four splint barcodes (*27*). **(C)** 96 sub-barcodes were used in each round of the split-pool procedure for barcode generation. Shown in a box-plot is the percentage of sub-barcodes incorporated to DNA barcodes at each position. The red line at 1.04% denotes the expected proportion (1/96). **(D)** Based on amplification and sequencing of UMIs associated with barcodes enzymatically released from individual beads, together with the fact that only ~50% of the barcodes have a dU at their stem and are expected to be cleaved, we estimate that each bead bears 491,811 (Interquartile range: 308,757 - 833,744) functionalized, barcoded oligonucleotides. **(E)** Schematic of a monolayer of hydrogel beads inside a 6% (v/v) PAA (polyacrylamide) encasing gel. **(F)** Differential interference contrast microscopy image of beads in an encasing gel. Yellow scale bar denotes 100 µm.



Supplementary Figure 2. Complementarity and expected outcomes of using twenty pairs of messenger sequence designs. (A) Fractions of oligos predicted to be bound at equilibrium in the messenger pool of forty sequences (20 functional messenger caps x 2 strands) at 37°C as calculated by Nupack (*28*). Twenty of these forty are complementary, resulting in high predicted fractions of hybridization. The orthogonality of the twenty pairs is shown by the symmetric pattern in the binding matrix. Complementary strands are adjacent in numbering, *e.g.* 1/2, 3/4 … 39/40. **(B)** We simulated the random assignment of a given number of bead subtypes to beads arrayed in a hexagonal lattice and computed, for any given bead, the proportion of immediately neighboring beads that are of the same subtype (*i.e.* homotypic neighbors).

Statistics of densely packed bead arrays

To validate our assumptions prior to attempting computational reconstruction, we collected basic statistics on barcoded bead arrays. With our protocol, we can reliably create surfaces of beads with 76% (s.d. 5.1%) coverage of the total area of a 2D plane, with an average bead diameter of 20 μm across the entire array (**Fig. 2A,B**). The packing constant for circles of equal size sets the maximum for this coverage at 91%, and our coverage could potentially be improved by decreasing variation in bead size. In hexagonal packing, the most dense way that equally-sized circles can be packed together, any given bead has six immediate neighbors. Using Delaunay triangulation on a representative, segmented image of the bead array, we find that the most common number of immediate neighbors for a bead is six (**Fig. 2C,D**).

SCOPE results in a spatially informative bead interaction count matrix that can be modeled

If SCOPE results in proximity-dependent reactions between DNA barcodes derived from physically adjacent beads, then the resulting count matrix should have a different distribution than a similar count matrix derived from performing SCOPE on beads in suspension. We found this to indeed be the case. After a SCOPE reaction on a 2D bead array, a bead's top 6 partners accounted for 66.7% (s.d. 17.4%) of interactions, and the top 20 for 93.7% (s.d. 9.1%). However, for a reaction in suspension, the 6 partners accounted for only 12.7% (s.d. 6.7%) of interactions, and the top 20 for only 35.0% (s.d. 16.3%) (**Fig. 2E,F**). Individual beads and interaction partners formed dense subgraphs akin to a tightly knit social network (**Fig. 2G**). An inverse square function closely modeled the ranked per-bead interaction count proportions (**Fig. 2H**). We designed a simulation around this inverse square framework to model a diffusive process of molecules between simulated beads arrayed in a hexagonal lattice. This simulation was flexibly parametrized with empirical per-bead distributions of total sent and total received messenger oligos.

Next, we trained a random forest regressor on simulated data (parametrized with empirical distributions) to convert sparse bead-bead interaction counts into pairwise distances. The values in this pairwise distance matrix are inversely related to the counts in the sparse interaction matrix, with "zeros" – pairs of beads with no detected interaction – assigned a single long-distance value. Although this approximation violates the triangle inequality at long distances, we posited that the dense local interactions between beads and their closest ring of neighbors might be best resolved with algorithms such as t-SNE or UMAP, which prioritize the preservation of local distances (*25*, *26*). This was in fact the case, as particularly for smaller numbers of simulated arrays, we found both t-SNE and UMAP were capable of accurately reconstructing local communities from a pairwise distance matrix inferred from simulated interaction counts (**Fig. S3A,B**).

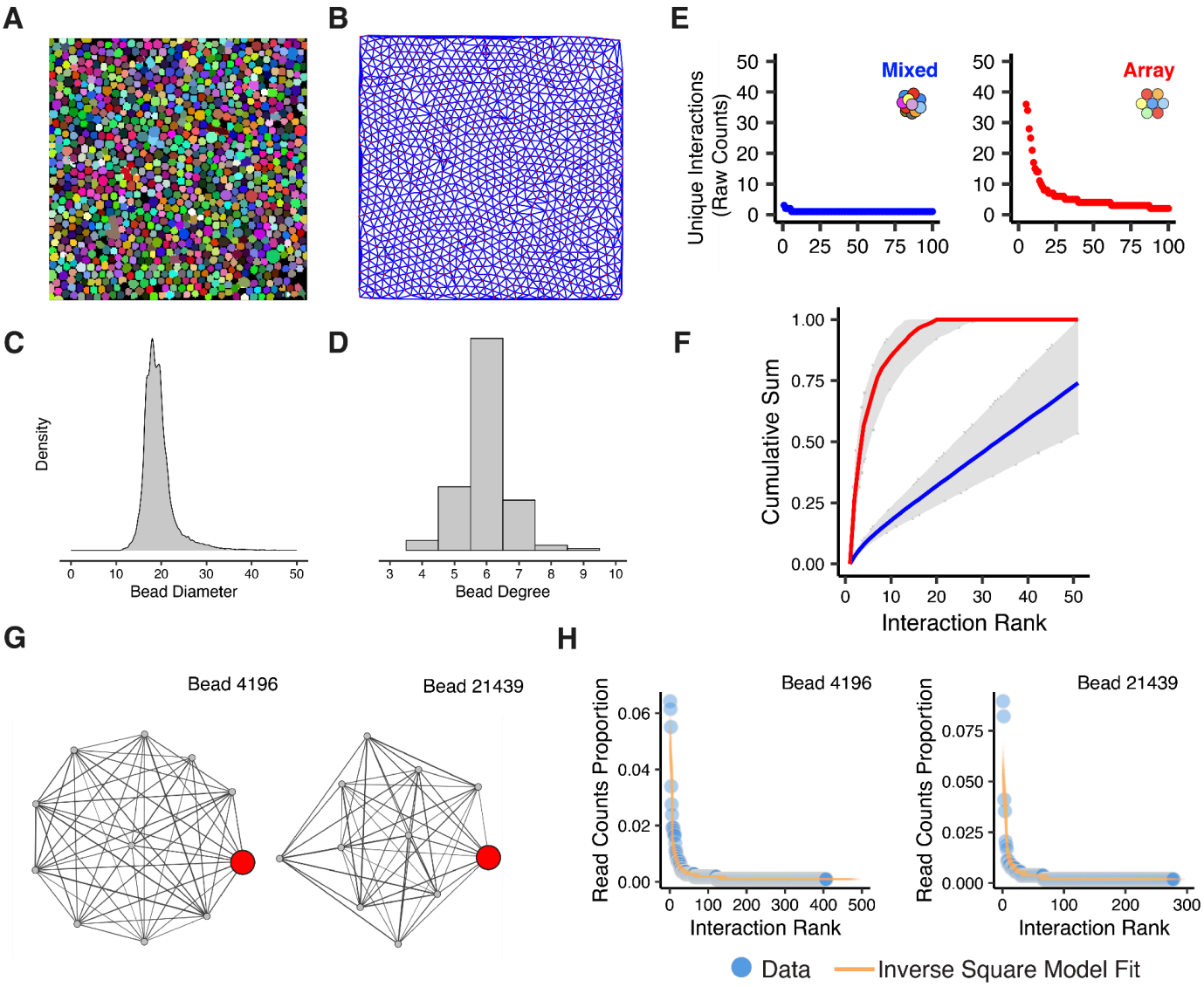
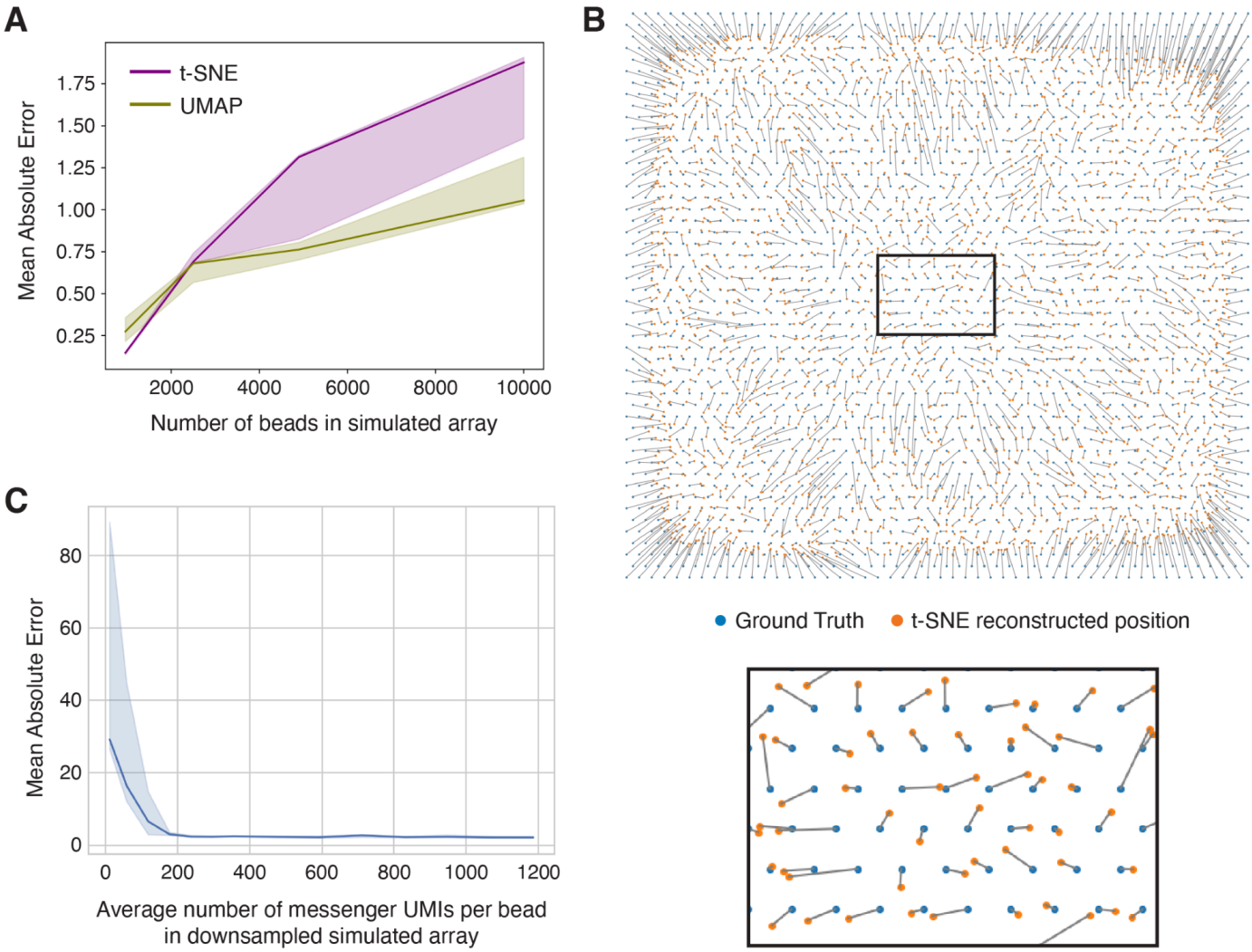


Figure 2. Bead arrays form local communities that can be reconstructed computationally. (A) Segmented mask of a representative image of a SCOPE bead array. **(B)** Density plot of bead diameters from the segmented image. **(C)** Delaunay triangulation for segmented bead centroids. **(D)** Distribution of the number of nearest neighbors observed for each bead. **(E)** Raw, unique interaction counts (*y*-axis) of a representative bead with all its interacting neighbors ranked in descending order (*x*-axis). Left (blue) plot shows counts for a representative bead from a SCOPE reaction performed in suspension. Right (red) plot shows counts for a representative bead from a SCOPE reaction performed in a 2D array. **(F)** Mean cumulative distribution function (CDF) of proportion of all interaction counts explained by the top neighbors of a given bead, for SCOPE reactions performed in suspension (blue) or 2D array (red). Gray shadings show the range of CDFs for the 1000 randomly sampled beads from which the mean CDFs were derived. **(G)** Network of interactions of two representative beads (red) from SCOPE reaction performed in 2D array with and among their top ten neighbors. **(H)** Proportion of read counts from interactions between the bead of interest and their neighbors are overlaid with 100 simulation instances from the inverse square model.



Supplementary Figure 3. Reconstructing simulated SCOPE bead arrays. (A) Mean absolute error of UMAP (green) or t-SNE (purple) applied to simulated SCOPE bead arrays of different sizes. Mean absolute error is measured as the difference in distance between an inferred position and the ground truth simulated position after a linear point cloud registration and point matching. The line shows the median of 5 trials and the shaded areas show the interquartile range. Units are in bead diameters. **(B)** A representative simulated array of 2500 beads reconstructed with t-SNE. Lines indicate matching between ground truth and t-SNE inferred positions after linear point cloud registration. Errors are greater at the boundary of the shape. **(C)** Mean absolute error of simulated and reconstructed arrays as a function of sequencing depth. Each point is the median of 10 trials. Shaded blue area shows the interquartile range for the set of trials.

Parallelized reconstruction and stitching of bead communities

Although t-SNE and UMAP were capable of reconstructing local communities consisting of fewer than ~2500 beads with reasonable accuracy, our procedure struggled to reconstruct large arrays, with accuracy tracking inversely with array size (**Fig. S3A,B**). Relying on the simulated dataset, we explored various approaches for solving larger arrays, ultimately settling on a dynamic programming approach in which we break a large array into a set of manageable sub-problems (**Fig. 3A**). Briefly, the global interaction matrix is first clustered and split into smaller, densely interacting sub-networks or communities. Given that we are most interested in accurate reconstructions at smaller scales, we chose to use t-SNE at a maximum cluster size of 2500 beads, a scale at which it outperforms UMAP (**Fig. S3A**). Next, boundary beads, which we define as those having a certain proportion of edge weights to beads from other clusters, are duplicated and assigned to both clusters. Due to the influence of the number of beads and sequencing depth on reconstruction quality (**Fig. S3C**), we found that there was no single value for tSNE's perplexity (a hyperparameter) well suited for reconstructing every cluster from a given dataset, let alone across datasets. Consequently, we opted to heuristically choose hyperparameters for each cluster that followed principles used by demographers when drawing political boundaries (**Fig. S4A-C**). First, cluster reconstructions at a given perplexity were required to be contiguous, *i.e.* resulting in only one "object" unified in space. Second, reconstructions must preserve graph topological structure, *i.e.* boundary beads must lie along the outer edge of the newly formed reconstruction and not at its center. Finally, each reconstruction must be compact, *i.e.* have points spread uniformly over the inferred area. When run in parallel, this hyperparameter search and t-SNE reconstruction allows for the rapid and robust inference of the local geography formed by each cluster of beads.

Because DNA barcodes are generated via a split-pool procedure, there is some risk of the same DNA barcode appearing in two beads, analogous to the collision problem encountered in single cell combinatorial indexing (*29*). Such collisions or 'doublet barcodes' risk confounding our spatial reconstructions by falsely drawing together two localities that are in fact at a distance (**Fig. S5A**). The number of such collisions is expected to increase as a function of the size of the array cast (**Fig. S5B**). To address this, we developed a clustering-based algorithm for detecting DNA barcodes associated with more than one true neighborhood in space. The resulting procedure can readily identify doublet barcodes with high precision and recall, both simulated and real datasets (**Fig. S5C-E**). Upon identification, these barcodes are removed from the reconstruction due to their ability to affect reconstruction accuracy, both locally and globally (*21*).

Once optimal reconstructions have been determined for each cluster of beads via t-SNE, we begin the process of stitching cluster reconstructions together to assemble a single reference frame. Starting from the cluster with the most boundary points, we add one cluster at a time, learning the optimal parameters of a rigid transformation (rotation, translation, and reflection) by minimizing the mean absolute error between the inferred positions of the boundary beads belonging to both cluster reconstructions through gradient descent (**Fig. 3A**; **Fig. S6**). The inferred positioning of all clusters is refined further by learning the parameters of a joint transformation, allowing a separate rigid transformation to be applied to each cluster until all boundary bead positioning errors are minimized. The result is a stitched and coherent global geometry, mapped from the formation of individual communities, deriving purely from local interactions between points on a 2D-plane.

Global inference of a map of randomly deposited barcoded beads with SCOPE

With this custom algorithm developed around simulated data, we next sought to apply it to SCOPE datasets generated experimentally. First, we generated a circular array of SCOPE beads with a diameter of 7 mm. As a

form of validation of the anticipated reconstruction, microliter droplets of polyA ssDNA oligonucleotides, complementary to the sender and receiver decoder oligos present on each bead, were hand-spotted on to the bead array in a predetermined pattern (3, 6 and 12 o'clock). Massively parallel sequencing of chimeras resulting from SCOPE, followed by application of the aforedescribed algorithm, resulted in the optics-free reconstruction of the spatial positions of 73,454 inferred beads (across 35 clusters) (**Fig. 3B,C**). Although our algorithm does not inherently constrain the shape that the global reconstruction will take, the pattern that we obtained was roughly a circle. Upon mapping reads corresponding to chimeras between manually spotted oligos and decoders, we obtained patterns consistent with expectation (**Fig. S7**). Of note, the 2D area occupied by 73,454 x 20 μm beads is about 60% of the area of a 7 mm circle, on par with our expected packing of 76%.

Next, we sought to evaluate whether we could apply SCOPE to reconstruct an arbitrary shape. For this, we fabricated an array of SCOPE beads cut into the shape of an asymmetric "swoosh" resembling the Nike logo (16.75 mm x 9.25 mm) (**Fig. 3D**). Massively parallel sequencing of chimeras resulting from the SCOPE reaction, followed by application of the aforedescribed algorithm, resulted in the optics-free reconstruction of the spatial positions of 60,477 inferred beads (across 32 clusters) which, encouragingly, resembled the Nike swoosh.

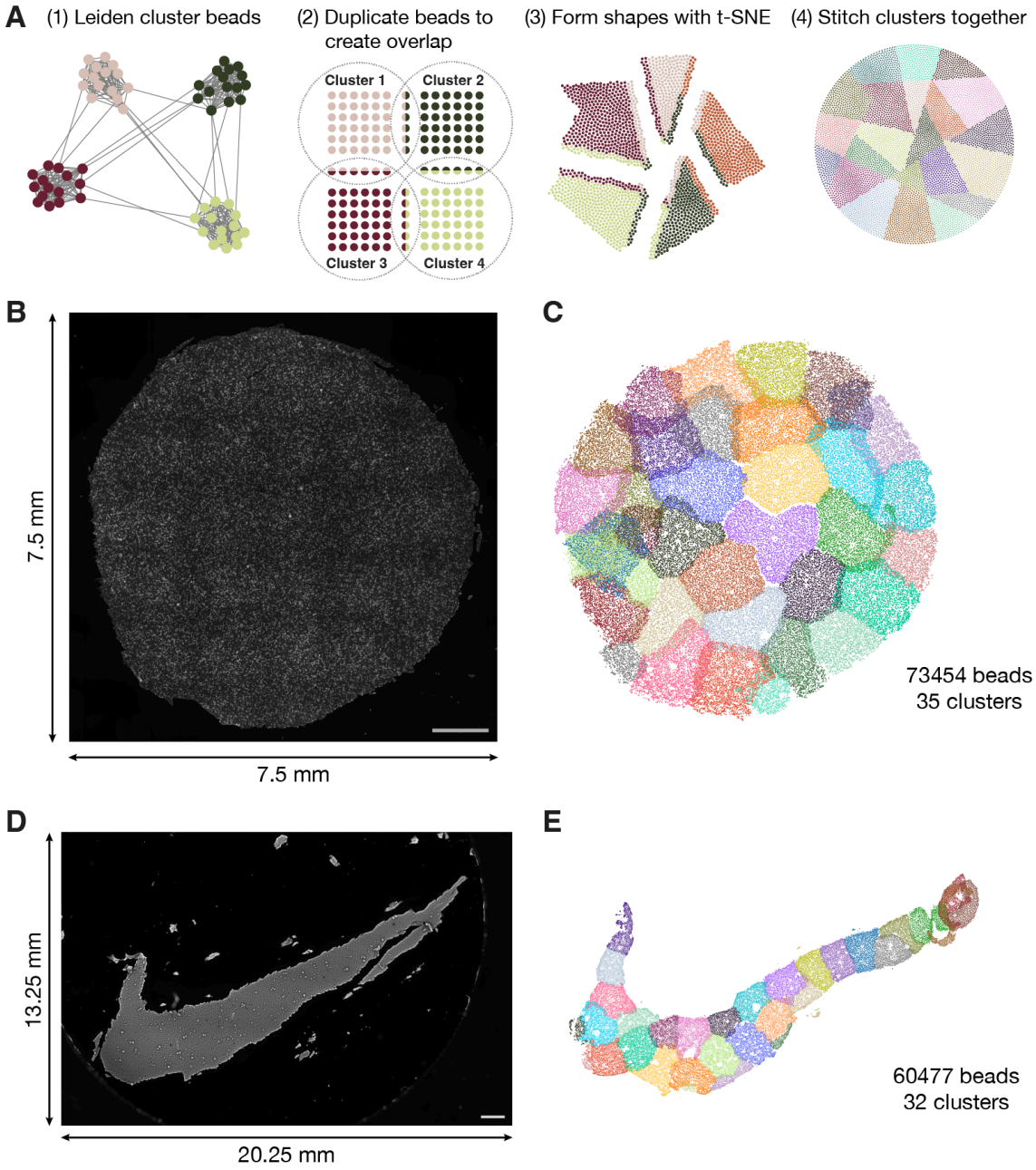
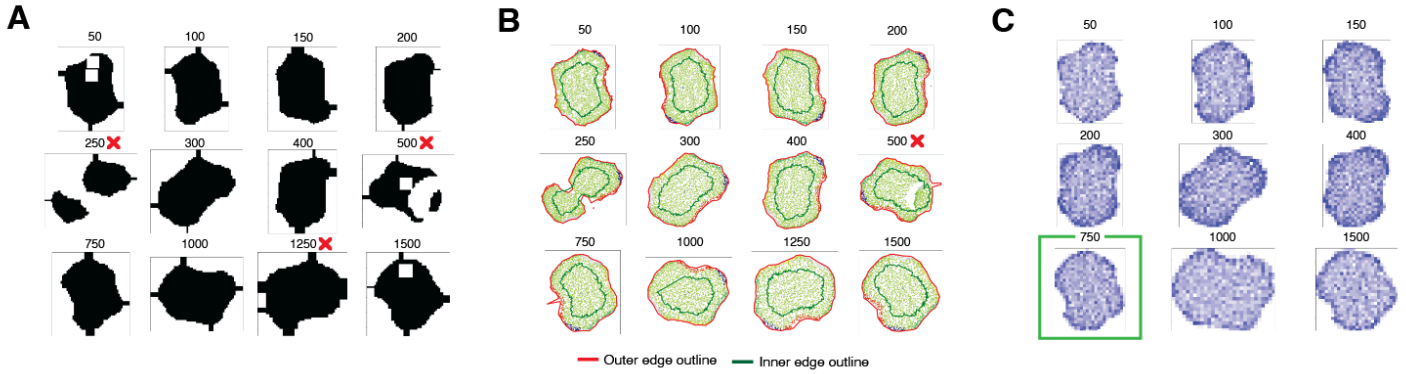
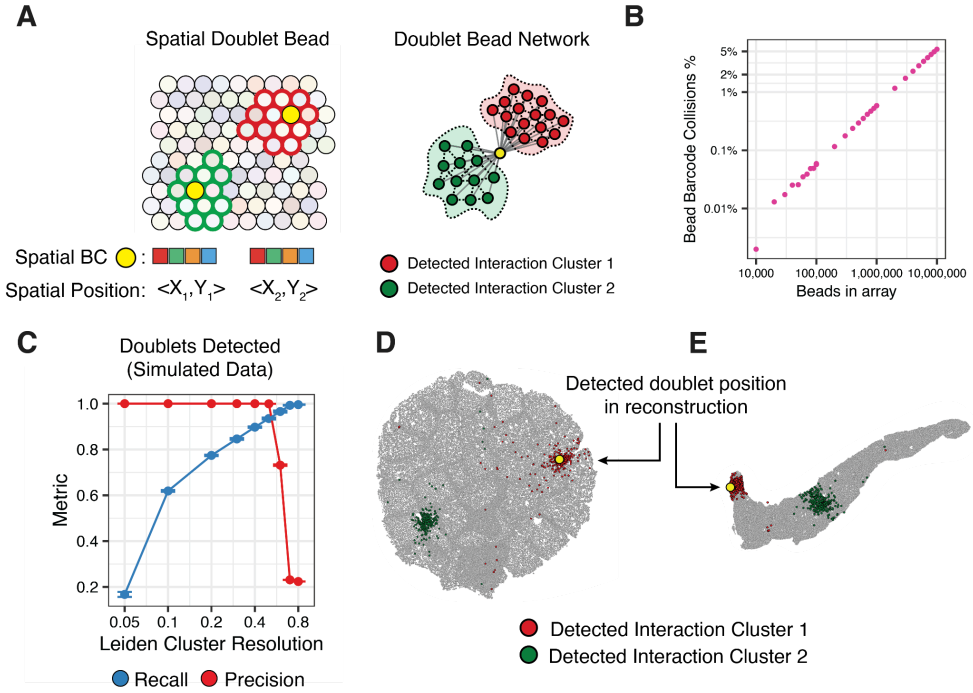


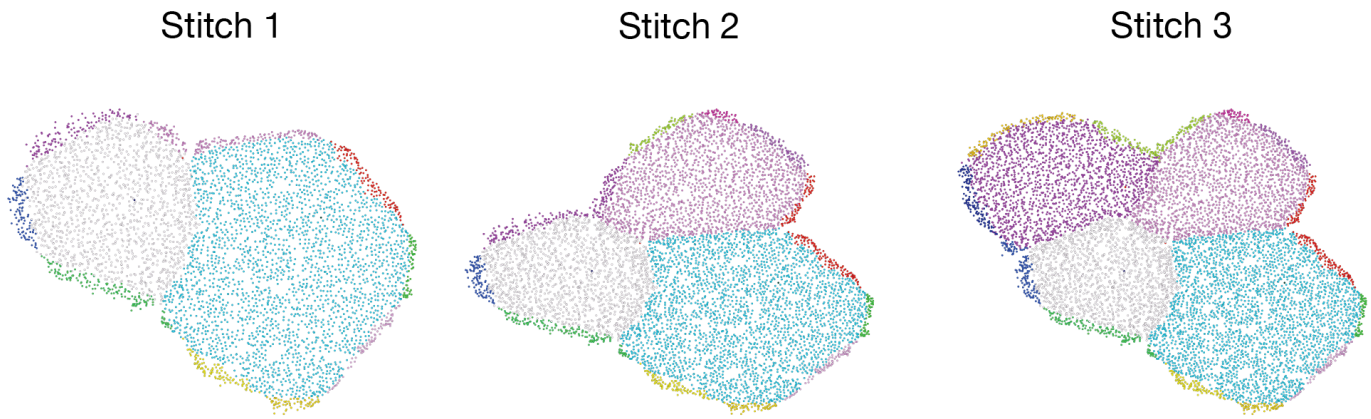
Figure 3. Optics-free reconstruction of 2D images via barcode proximity graphs. (A) Overview of the spatial reconstruction algorithm. **(B)** Single channel image of a manually cut circular bead array with 7 mm diameter. Scale bar denotes 1 mm. **(C)** Computational reconstruction of the circular bead array with SCOPE. **(D)** Stitched single channel image of an asymmetric bead array manually cut to “swoosh” shape and stained with SYBR Gold. Scale bar denotes 1 mm. **(E)** Computational reconstruction of the asymmetric bead array with SCOPE. The reconstruction coordinates have been rotated to match the microscopy image shown in panel **D**.



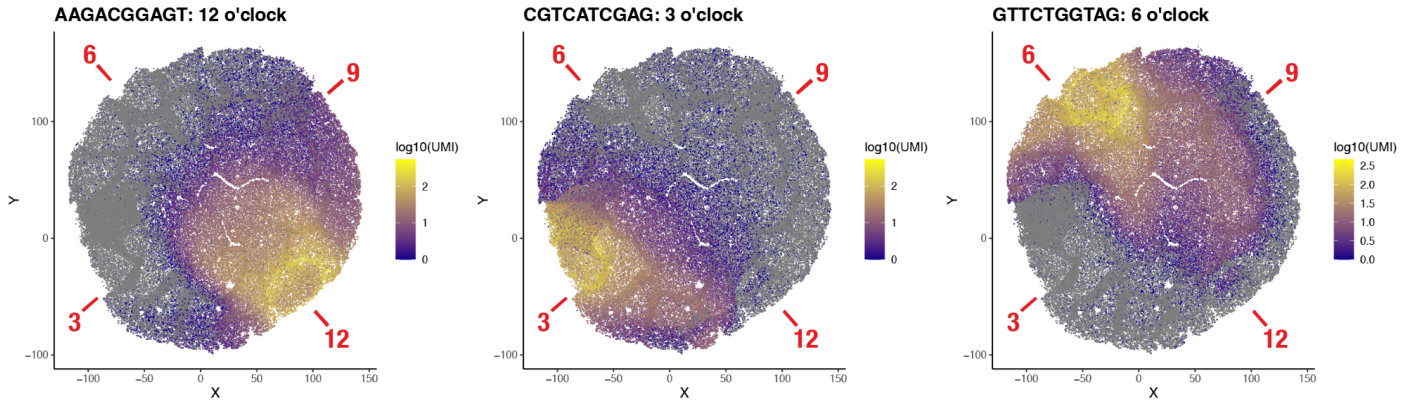
Supplementary Figure 4. Acceptance criteria for valid local reconstructions. (A) A binary mask of the cluster at various t-SNE perplexity values, shown above each reconstructed cluster after binary erosion to remove noise. X's (red) indicate rejected solutions (those that are in more than one piece) and the lack of any identifier indicates a passing solution. **(B)** The alpha shape containing each cluster (in red) is shown in relation to the shape bounding the majority of the non-boundary points (green). X's (red) indicate rejected solutions and the lack of any identifier indicates a passing solution. **(C)** Plots depicting the 2D density of points within a cluster faceted by various values for the perplexity (shown above each reconstructed piece). The green box indicates the chosen perplexity with the most uniform distribution of points after the filtering steps depicted in panels **A** and **B**.



Supplementary Figure 5. Doublet detection by community detection. (A) Doublets are defined as beads that have two spatial positions but share an identical DNA barcode (ground truth – left). In sequencing data these doublets appear as a single bead with collapsed interaction networks. **(B)** Expected rates of doublet occurrence for arrays of a given size. Estimated by computing the birthday problem for a range of array sizes with a fixed set of 96^4 barcodes. **(C)** Doublets detected in simulation using leiden clustering and a range of cluster resolutions. Recall and precision are shown with error bars indicating standard deviation. **(D,E)** Examples of detected doublets in reconstructed data. Highlighted in red and green are the two communities detected for a single bead (shown in yellow).



Supplementary Figure 6. Serial stitching procedure. Serial stitching proceeds by aligning boundary beads shared between a new piece and the working solution using gradient descent to learn rigid transformations. Pieces that share the most boundary beads with the working solution are added iteratively until all pieces are exhausted.



Supplementary Figure 7. Predetermined pattern of spotted poly-dA hash oligos informs orientation of symmetric arrays. Decoder molecules capture barcoded poly-dA tailed hash oligos that were manually spotted onto the circular bead array at the 12 o'clock, 3 o'clock, and 6 o'clock positions, denoted by the red ticks and labels (reconstruction output coordinates have not been rotated). The three different barcoded hash oligos are locally concentrated at the expected positions, which can inform what rotation or reflection is needed to match the original bead array if the bead array is symmetric.

An eye examination for an optics-free SCOPE reconstruction

The resolution of optical systems is limited by the wavelength of light used, as well as the effective numerical aperture of a microscope's optical components. In principle, DNA-based spatial reconstructions are not subject to the same physical constraints as optical systems, but rather by the parameters of diffusion kinetics. Although we are far from approaching the limits of microscopy in these experiments, we sought to evaluate SCOPE's ability to reconstruct images while preserving scaling at varying levels of resolution, as well as to resolve multiple features ("colors") via decoder-based capture. To this end, we performed a virtual eye examination by attempting to use SCOPE to reconstruct aspects of the classic Snellen eye chart for visual acuity. Using a microarray printer, we deposited picoliter droplets of 13 polyA-tailed oligonucleotide "paints" with a spot size of 100 μm and spot-to-spot center pitch of 50 μm in the form of a Snellen eye-exam chart, directly onto a 2D SCOPE array. Three of the 13 oligonucleotides bore a dye allowing for their direct visualization via microscopy (**Fig. 4A**). The size of the resulting spotted array was 17.18 mm x 40.97 mm and contained striated patterns corresponding to a regular shift based on the number of spots needed to print each line. Based on imaging, we estimate that each printed oligonucleotide spot covered an average of 45 beads (s.d. 1 bead) in the cast array.

The Snellen chart is designed to test visual acuity of the eye and contains a series of letters arranged from largest to smallest, *i.e.* least to most challenging to resolve. For SCOPE reconstructions, the degree of difficulty is arguably inverted. Due to the reliance on purely local interactions for reconstruction, patterns spanning large areas require assembly of individual features followed by stepwise stitching (*e.g.* large "E" – **Fig. 4A, top right**), while small letters that fit within a single cluster require only a single local reconstruction step (*e.g.* small "D" – **Fig. 4A, bottom right**). Following the SCOPE reaction, we generated two sequencing libraries: the first amplified from messenger chimeras capturing information about bead-bead interactions, and the second amplified from decoder chimeras capturing information about the positions of microarray-spotted polyA-tailed oligonucleotides. Massively parallel sequencing of the resulting SCOPE messenger sequencing library identified DNA barcodes corresponding to 1.40 million beads collectively engaging in 193 million unique bead-bead interactions. Sequencing of the accompanying decoder library linked an average of 53 (s.d. 67) polyA-tailed "painted" molecules to each bead.

Because our current heuristic for "stitching" encountered error-propagation challenges when applied to 582 neighborhoods comprising the entire Snellen eye chart in one shot (**Fig. S8A**), we leveraged an initial global reconstruction to identify subsets of communities corresponding to the letters "E" (first row), "F" (second row), "O" (third row), and "D" (fourth row) (**Fig. 4A; Fig. S8B**). The beads derived from each of these four subsets were then subjected to the full reconstruction algorithm (**Fig. 3A**), resulting in the grossly accurate reconstruction of each individual letter (**Fig. 4B,C**).

For a more formal quality assessment, we focused on the large letter E, which was the most challenging with respect to the number of beads (n = 182,247) and clusters (n = 82) inputted to stitching. During reconstruction, SCOPE only has access to the messengers and cannot access the decoders. Therefore, we were encouraged that a clear striated pattern was evident in the reconstruction when each decoder-library-derived "oligonucleotide channel" was visualized individually (**Fig. 4C-D**). To compare the dimensional scales of the original image vs. its optics-free, SCOPE-based reconstruction, we measured the lengths of various features of the "Opticon Sans" styled letter E, normalized to the letter's baseline, and found them to be highly correlated (r = 0.978, Pearson's ρ) (**Fig. 4E**). This result suggests that SCOPE reconstructions can recover local and global length scales ranging from the centimeter scale to the micron scale, while preserving the relative length scales of the original image.

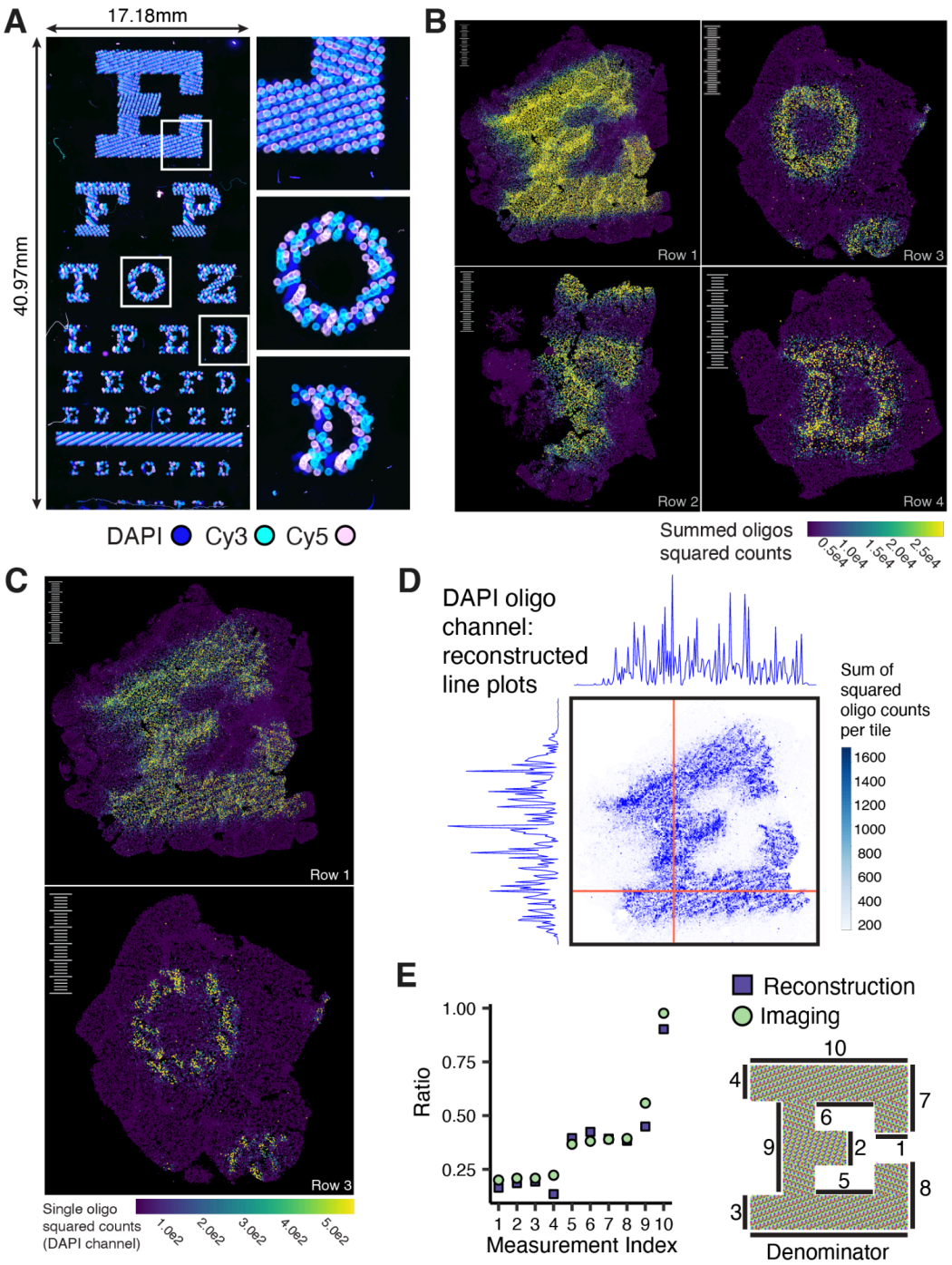
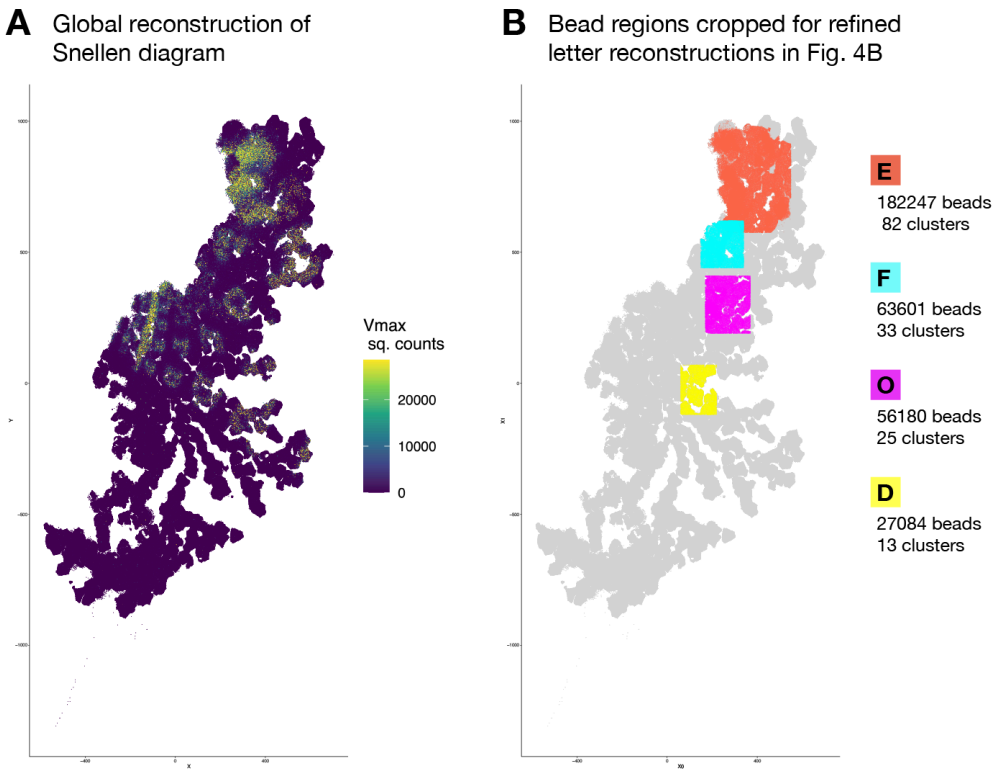


Figure 4. An eye examination for SCOPE reconstructions. (A) Composite fluorescent image of a Snellen eye chart for visual acuity printed onto a SCOPE array using a microarray printer. Of the thirteen printed oligonucleotides, three contained fluorescent dyes either in solution (DAPI) or conjugated to the oligo directly (Cy3, Cy5). Insets of the "E" (top), "O" (middle), and "D" (bottom) are shown for reference. **(B)** Maximum oligonucleotide projections of select refined reconstructions. The letters "E", "F", "O" and "D" originating from the 1st, 2nd, 3rd and 4th rows, respectively, are shown. The reconstructions of each letter are shown at different scales, but with a common ruler marker. **(C)** Maximum intensity projection of a single oligo channel (DAPI channel oligo) from the "E" in Row 1 and the "O" in Row 3. **(D)** Locally binned image colored by the number of oligos in each binned pixel. Line intensity profiles depicted for the horizontal (trace shown above) and vertical (shown on left) red lines crossing the binned image. (**E**) Ratio of measurements between the bottom margin, *i.e.* baseline, of the letter E and each margin indicated in the inset (right).



Supplementary Figure 8. Global reconstruction of Snellen eye chart. **(A)** Attempt to globally reconstruct the entire Snellen diagram in one shot, with beads colored by maximum projection of captured oligonucleotide squared counts. From the sequencing data used for this reconstruction, we detected 1,395,536 beads, which were divided among 582 clusters for local t-SNE based reconstruction followed by stitching. The low quality of the global reconstruction appeared primarily due to error propagation at the cluster stitching level. The relatively low depth at which this particular library was sequenced may have also contributed. **(B)** Bead regions of interest were chosen for smaller scale reconstruction to refine the visualization of the printed letters. Deeper sequencing of the messenger chimera libraries was performed prior to attempting reconstructions of individual letters.

Discussion

The paradigm of molecular self-assembly is deeply encoded in the evolutionary histories of molecules, cells and organisms (*30*). Biological systems are impressive not only for the information density encoded by their genomes, but also for the scales over which this information is capable of propagating, *e.g.* a micron-scale zygotic nucleus encoding a 30 meter, 200 ton blue whale. Molecular and cellular interactions mostly occur at a local scale because they are often diffusion-limited and rely on built-in circuitry to spatially propagate effects. Inspired by these design principles and bolstered by the simple physical rules that govern the interactions between DNA molecules, we developed SCOPE, a method that leverages local interactions to enable self-registration and molecular capture spanning four orders of magnitude in length scale.

SCOPE relies on barcoded hydrogel beads that are able to send information regarding their identity to neighboring beads through highly localized, diffusion-limited DNA-DNA proximity reactions. We demonstrate that this form of pairwise interaction data can be computationally transformed and deconvoluted to reconstruct 2D images of arbitrary size, shape and length-scale, provided adequate sampling. A key point is that SCOPE reactions take place in *constant experimental time* regardless of the size of the array. We foresee that self-registering arrays could be broadly useful in spatial genomics, where current state-of-the-art methods rely heavily on microscopy either for the registration of barcode sequences to a spatial position or for the registration of molecules within biologically intact samples. This reliance on optical systems creates an inherent trade-off between a sample's physical size and the optical resolution of the system. These challenges cast the collection of large datasets spanning many 2D sections (*e.g.* whole mouse brain (*31*, *32*)) as feats of organization as opposed to routine experiments. The extension of SCOPE to capture and decode the relative spatial positions of biomolecules or cells originating from biologically intact samples is thus a logical next step.

Experimentally and computationally, what limits SCOPE? Barcoded beads can be produced at exponential scales with split-pool strategies, and are easily cast to generate large 2D arrays. As noted above, the SCOPE reaction takes place in constant experimental time regardless of array size, and even for very large SCOPE arrays, reaction volumes are trivial. However, particularly for large arrays such as the Snellen eye chart and beyond, the depth of sequencing of chimeric sender-receiver messenger libraries remains as a potentially limiting factor. Algorithmically, SCOPE reconstructions utilize graphical methods that can be parallelized over many processors. Our current algorithm can yield faithful solutions of arrays consisting of ~80 clusters (~200,000 beads); however, global reconstructions of larger scales (*e.g.* the entire Snellen eye exam chart) breaks down due to error propagation at the cluster stitching level. We anticipate that the issues can be solved through the application of graph drawing algorithms (*33*), by incorporating a one-step global reconstruction method as an initialization step (*22*) and/or by deeper sequencing of chimeric messenger libraries.

Experimental molecular and cellular biology increasingly leverage the concepts of DNA barcodes (*e.g.* UMIs, MPRAs, combinatorial indexing) and spatial anchoring (*e.g.* massively parallel sequencing) in order to "parallelize" experimentation within single reaction volumes. SCOPE brings these concepts together in one particular form to decode 2D images in an optics-free manner. Looking forward, we envision that more sophisticated intersections of DNA barcoding and spatial anchoring, *in vitro* and perhaps *in vivo*, will be fruitful, *e.g.* for molecular computing, for encoding and decoding enormous numbers of experiments, for exchanging information between cells during biological development, or for building spatially coherent DNA automata.

Methods

Molecular biology methods

Hydrogel bead fabrication

Hydrogel beads were produced through the use of a flow focusing microfluidic device to create water in oil emulsions (*34*). First, acrydite-modified oligos were designed and ordered from IDT with the TruSeqR1 (OLG_001: /5Acryd/TTTTTTT/ideoxyU/CTACACGACGCTCTTCCGATCT) and TruSeqR2 (OLG_002: /5Acryd/TTTTTTTTTTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT) Illumina sequences. Acrydite oligos were stored at -20°C. The TruSeqR1 handle contained a deoxyuracil for USER enzyme-controlled cleavage off the bead. To make the polyacrylamide gel beads, a DNA and acrylamide mix (3% (v/v) acrylamide/bis solution (Sigma), 3% acrylamide solution (Sigma), 48 mM Tris–HCl pH 8.0, 0.25% (w/v) ammonium persulfate, 0.1X Tris-buffered saline–EDTA (TBSET: 10 mM Tris–HCL pH 8.0, 137 mM NaCl, 20 mM EDTA, 1.4 mM KCl, 0.1% (v/v) Triton-X 100), 50 µM of OLG001, and 50 µM of OLG002) was first prepared. The TEMED (Bio-Rad) catalyst was held out of the aqueous phase to prevent premature polymerization and added into the emulsion collection tube instead. The acrylamide mix was run through a droplet generator (Droplet Genomics) set up with a DG-DM-25 chip (Atrandi Biosciences) as the aqueous phase at 200 µL/hr, along with 2% RAN-008-FS v/v (Ran Biotechnologies) in HFE7500 (Oakwood Chemical) as the oil phase at 300 µL/hr. The resulting emulsions were allowed to polymerize overnight at room temperature. To break the emulsions, 150 uL of 1H,1H,2H,2H-Perfluorooctanol, 97% (Thermo Scientific Chemicals) was added for every 1 mL of beads produced. The bead solution was vortexed and centrifuged for 1 minute at 1000 x g and excess oil was removed from the polyacrylamide bead layer. Two more washes with a 1:1 ratio of hexane to TBSET were performed to further remove residual oil. Finally, multiple TBSET washes were performed until all the residual oil was removed.

Hydrogel bead barcoding and functionalization

Four consecutive rounds of splint barcodes were ligated together to build the bead barcodes. Eight plates of top and bottom oligos were ordered for the four barcode splints (**Tables S1-S4**). The following protocol closely followed the Delley & Abate protocol to make combinatorially barcoded hydrogel beads (*27*). Briefly, polyacrylamide beads were prepared for splint ligation by first annealing primer OLG_003 (**Table S6**) to the polyacrylamide-incorporated DNA stub on the bead. This step creates a four base pair overhang as a handle on which to ligate the first barcoded splint. Barcoded splints were ordered as two separate oligos to form the top and bottom portions of the splint and mixed at a 1:1 ratio. If the splint components were ordered without a 5' phosphorylation modification, T4 PNK (NEB) was used to add phosphate groups to the top and bottom oligos before each round of ligation. For the first round of barcoding, beads were distributed across a 96 well plate containing barcoded splints and ligated with T4 ligase (NEB). The plate was placed in a thermomixer at 37°C for 2 hours, shaking at 1000 rpm to prevent beads settling at the bottom of the wells. The T4 ligase reaction was inactivated at 65°C for 20 minutes before the beads from each well were pooled together and mixed well. The barcoding and split-pooling steps were repeated for a total of four times to create a 96^4 combinatorial barcode space.

Once the beads have been uniquely barcoded via splint ligation, a capping sequence was added to the end of the barcode to functionalize the bead oligo. Since the bead proximity reactions rely on creating 20 subtypes of beads that can interact with beads of another subtype but not with their own, the beads were divided into 20 pools for 20 separate functionalization reactions. Each functionalization reaction contains 1.5 parts of each of the 19 antisense capping oligos and 20 parts of the single sense capping oligo (**Table S5**;

Sense_messenger_001–020, Antisense_messenger_001-020). For example, one of the reactions would contain 20 µL of Sense_messenger_001 and 1.5 µL each of Antisense_messenger_002 through Antisense_messenger_020 (Antisense_messenger_001 was left out). The entire mixture of 19 antisense to one sense oligos was then mixed with a poly-dT capping oligo (**Table S5**; Decoder_001) at a 1:1 ratio, and then ligated onto the fourth bead barcode splint with T4 ligase. Thus, both decoder and messenger species were loaded onto the bead barcodes and the beads were able to perform poly-A molecule capture and bead-bead proximity interactions, respectively. Finally, to get rid of incomplete DNA oligo stubs resulting from inefficiencies at any of the previous ligation steps, beads were treated with ExoI nuclease after annealing complementary oligos to all possible functional 3' caps. Beads were then stored at 4°C for up to one year.

Casting densely packed monolayers of barcoded beads immobilized on glass slides

To immobilize barcoded beads in a dense monolayer, 15 µL of packed beads were first mixed with 2.75 µL of Bis-acrylamide 37.5:1 (Bio-Rad) to form a bead slurry. 0.55 µL of freshly prepared 5% (w/v) APS (Bio-Rad) was added to the bead slurry and the entire solution was pipetted as a large droplet onto a 25.4 mm x 76.2 mm microscope glass slide functionalized with 3-(Trimethoxysilyl)propyl methacrylate (Sigma-Aldrich) as described previously (*19*). A 22 mm x 22 mm coverslip was placed over the bead slurry and lightly pressed down to spread the beads into a single layer. These reagent volumes were scaled for casting larger arrays. The spreading of the beads was checked by eye under the microscope to ensure a densely packed monolayer under the area of the coverslip.

Once beads have been spread underneath the coverslip in the acrylamide mixture, we sought to polymerize the acrylamide directly onto the functionalized glass surface. As oxygen inhibits the polymerization reaction, a chamber was prepared using a one-gallon Ziploc bag filled with inert argon gas to flush out the oxygen or set in an anaerobic chamber (Coy Labs). To catalyze the polymerization of the encasing gel, 2% (v/v) TEMED solution was pipetted around the edges of the coverslip and the slide was placed in an anoxic chamber overnight. Once the gel had polymerized in the glass slide was incubated in a tray with enough water to cover the slide for at least 10 minutes at room temperature. After the encasing gel had been hydrated, a new razor blade was inserted under the edge of the coverslip and used to peel the coverslip away from the bead array. 1 mL of water was then used to rinse the bead monolayer for a total of five times, and the rinsed slide was placed into a tray with enough water to cover the monolayer for 10 minutes. These rinsing steps were for washing off unpolymerized acrylamide monomers in the encasing gel that inhibit subsequent PCR reactions. The bead monolayer was then dried uncovered in the fume hood and stored at 4°C until use.

Printing images on bead arrays with oligonucleotide inks

Thirteen poly-dA oligos (OLG_007 through OLG_019) were first reconstituted to 100 µM in IDTE pH 7.5 and then diluted to 1 µM in printing buffer (0.01% Tween-20 (v/v) and 0.5% (v/v) glycerol). Because the FAM-conjugated poly-dA oligo displayed weak fluorescent signal during imaging, DAPI was included into the printing buffer for this particular oligo so that those spots emitted fluorescence in the DAPI channel. Each oligo was loaded into a separate well of a 384 well-plate (Scienion CPG-5502-1) and spotted using the Scienion Sciflexarrayer S3 using a piezo dispense capillary 100 (PDC 100). A custom map was loaded for each oligo and prints were conducted serially with a spot-to-spot spacing of 100 µm with a pitch of 50 µm. After printing, the bead array was imaged using the Keyence BZ800 with a 4x objective. Tiled images were stitched and saved for image processing.

SCOPE reactions for messengers and decoders (circular and asymmetric arrays – Fig. 3)

The desired shape was made in the dried bead monolayer by scraping the excess beads off the glass slide with a clean razor blade. To constrain poly-dA hashing oligos to specific areas of the bead monolayer for capture, 0.2 µL of 0.1 µM poly-dA hashing oligo (**Table S6**; OLG_004 through OLG_006) was pipetted directly onto the bead monolayer and let dry in the fume hood. Next, a silicone chamber (Grace Bio-Labs, 20 mm diameter, 2.5 mm height) with an adhesive back was adhered to enclose the shaped bead monolayer. For the USER and reverse transcription reaction, a master mix of 1X CutSmart (NEB), 1X SuperScript IV RT Buffer (Thermofisher), 500 µM dNTPs (NEB), 67 U/mL of USER enzyme (NEB), 10000 U/mL of SuperScript RT Enzyme (Thermofisher) was prepared. For a 20 mm diameter chamber, 50 µL of USER and reverse transcription master mix was added to the center of the bead monolayer. An 18 mm diameter glass cover slip round was gently placed on top of the bead monolayer so that the master mix is spread evenly underneath the coverslip and covers the entire area of the monolayer.

The slide was then placed in a pre-warmed thermal cycler with an adaptor that can directly transfer heat evenly from the thermal cycler block to a glass slide. Care was taken to ensure that the lid of the thermal cycler does not press down on the slide with the adherent chamber, with PCR tubes placed at the edges of the thermal cycler block to keep the thermal cycler lid off the slide. The glass slide was then incubated at 37°C for 15 minutes and 55°C for 15 minutes for the USER and reverse transcriptase reactions, respectively. Once this USER and reverse transcription reaction was finished, the slide was taken out of the thermal cycler and a clean razor blade was used to carefully remove the coverslip on top of the bead monolayer. To collect the USER-cleaved barcodes from the bead monolayer, which form the sender-decoder and proximal poly-dA oligo chimera, 200 µL of water was added into the chamber and a P200 pipette was used to triturate the solution gently. The entire volume of solution inside the chamber was collected into a new 1.5 mL tube, which was then used for generating the library of poly-dA molecules that have been captured by a spatial bead barcode.

To collect bead interaction information, which is based on sequencing counts of sender-messenger and receiver-messenger chimeras, the bead monolayer in the chamber was washed three more times with 300 µL of water. After removal of the last wash, 100 µL of 0.1 N NaOH was added into the chamber, covering the bead monolayer. The solution was incubated at room temperature for 15 minutes to denature the sender-receiver chimeric molecules off of the beads. Then, the entire volume of the chamber was collected into a new 1.5 mL tube. 100 µL of 0.1 M Tris-HCl pH 8.0 was pipetted into the chamber and triturated gently several times with a P200 pipette to neutralize the previous NaOH wash and rinse the remaining denatured sender-receiver chimeric molecules from the bead monolayer. The entire volume was then collected from the chamber and added to the 1.5mL tube with the previously collected NaOH wash.

A SPRI reaction was performed on the entire poly-dA collection volume to size select for the decoder library and eluted in 20 µL of water. On the entire volume of denatured sender-receiver messengers, a SPRI cleanup was performed and eluted in 20 µL of water. These two elutions were then prepared as Illumina sequencing libraries via an indexing PCR step to add the Illumina handles and sequencing indices.

Indexing PCRs for SCOPE libraries (circular and asymmetric arrays – Fig. 3)

A 30 µL indexing PCR reaction was prepared with 15 µL of NEBNext High-Fidelity 2X Master Mix (NEB), 10 µL of template DNA (SPRI purified from the previous section), and 4 µM each of forward and reverse indexing primers (**Table S6**). Messenger reactions used TruseqP5 (OLG_025) and Truseq P7 (OLG_026) indexing primers, while decoder reactions used NexteraP5 (OLG_028) and TruseqP7 (OLG_026) indexing primers. SYBR Green was added to track the number of PCR cycles before saturation and the PCR was performed with

an annealing temperature of 69°C and an extension time of 45 seconds. The PCR reaction was stopped before the qPCR SYBR Green curve went past the exponential phase. A final SPRI cleanup was performed on the crude PCR product from the decoder-polyA molecule capture library, as well as a final SPRI cleanup of the crude PCR product from the sender-receiver messenger library. The concentration of the libraries was quantified using a High Sensitivity D1000 ScreenTape on a TapeStation 4200 system. Most sequencing runs were performed with P2 200 cycle kits on NextSeq 2000 (Illumina) with standard chemistry. Reads were paired-end spanning 105 base pairs for Read1, 105 base pairs for Read2, and 10 base pairs for both P5 and P7 indices. This read structure provides two times the sequencing coverage over the portion where annealing happens in the chimeric sender-receiver molecule. The same read structure can also accommodate the decoder-captured poly-dA molecule library, so that it can be sequenced on the same run as the messenger library. To increase sequence diversity, between 10 and 15% Phi-X was spiked into the sequencing library and runs were loaded at 850 pM.

SCOPE reactions for messengers and decoders (large eye exam array - Fig. 4)

Since the oligo spots that were printed onto the bead array show outwards diffusion as in **Fig. S7**, we sought to limit the diffusion of these printed oligos so that letter shapes with sharp boundaries can be recovered. After the oligos were printed on the bead array, three poly-dA oligos (OLG_020 through OLG_022) were diluted to 0.1 μM and a single oligo was used in consecutive washes of the array. Thus, we could soak up the available decoder molecules on the beads that were not in contact with the printed oligo spot through hybridization with these background poly-dA oligos.

Given the larger size of the bead array for the eye exam diagram reconstruction compared to the previous circular and asymmetric arrays, the reaction volumes were scaled accordingly based on the fold change in surface area of the arrays. A glass coverslip was placed on top of the bead array during the two-phase USER and reverse transcriptase reaction, as described for the circular and asymmetric arrays. To collect the supernatant containing the sender-decoder and poly-dA oligo chimeras, the array was washed with a low salt buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 3 mM MgCl2, 0.1% (v/v) Tween-20, 0.1% (v/v) NP-40). The messenger library was collected by incubating the array in 300 μL of 0.2 N NaOH at room temperature for 7 minutes and quenching with 50 μL of 1 M Tris-HCl pH 8. The collected decoder and messenger libraries were then SPRI purified as described for the circular and asymmetric arrays.

Indexing PCRs for SCOPE libraries (large eye exam array – Fig. 4)

For the decoder library with captured poly-dA hashes, a 200 μL indexing PCR reaction was prepared with 100 μL of NEBNext High-Fidelity 2X Master Mix (NEB), 90 μL of template DNA (SPRI purified from the previous section), and 0.2 μM each of NexteraP5 forward (OLG_028) and TruseqP7 reverse (OLG_026) indexing primers (**Table S6**). SYBR Green was added to track the number of PCR cycles before saturation and the PCR was performed with an annealing temperature of 60°C and an extension time of 30 seconds. The PCR reaction was stopped at 6 cycles. A 0.8X SPRI cleanup was performed on the crude PCR product from the decoder library to generate the final sequencing library.

For the messenger reactions containing bead proximity information, a 300 μL PCR reaction was prepared with 150 μL of NEBNext High-Fidelity 2X Master Mix (NEB), 100 μL of SPRI purified messenger products, and 0.2 μM each of TruseqP5 forward (OLG_023) and TruseqP7 reverse (OLG_024) primers (**Table S6**). These primers did not contain a sequencing index, which would be later added in the subsequent PCR reaction. SYBR Green was added to track the number of PCR cycles before saturation and the PCR was performed with an annealing temperature of 63°C and an extension time of 30 seconds. The PCR reaction was stopped at 8

cycles. Two consecutive 0.85X SPRI cleanups were performed. The SPRI purified template was then put into a 100 µL indexing PCR reaction with 50 µL of NEBNext High-Fidelity 2X Master Mix (NEB) and 0.2 µM each of TruseqP5 forward indexing (OLG_025) and TruseqP7 reverse indexing (OLG_026) primers (**Table S6**). Both libraries were sequenced on the Illumina Nextseq 2000 as previously described for the circular and asymmetric arrays.

Computational methods

Fastq read processing for bead-bead interaction matrix

From the fastq file of sequencing the messenger library, the four 10-bp sub-barcodes of the combinatorial spatial barcode were extracted from Read1 and error corrected to known barcodes through the following process. In our final bead barcode, known 4-bp "scars" exist between each of the sub-barcodes from ligating the barcoded splints together. Additionally, the set of 96 sub-barcodes at the first position contain extensions between 0 and 3-bp, which need to be determined to extract the rest of the sub-barcodes. To do so, each pair of sequencing reads was first mapped to scar sequences to determine the extension length of the first of the four sub-barcodes (between 0 and 3-bp), as well as the positions of each sub-barcode. Next, the four extracted sub-barcodes were mapped to true sub-barcode sequences. The annealed portion between the sense and antisense ends of two bead barcodes were also mapped to known messenger sequences.

For error correction of the sequences, certain numbers of substitutions were allowed to account for sequencing errors via calculating the Levenshtein distances between them. The Levenshtein distance between two sequences is the minimum number of single-base edits (insertions, deletions or substitutions) required to change one sequence into the other. For scars mapping, we tried different extension lengths and took the extension length that minimizes the Levenshtein distance and then allowed the Levenshtein distance to be lower than 2. For barcode mapping, we allowed each of the 4 barcodes to have a Levenshtein distance lower than 2. For mapping of the annealed region between the two barcodes, we corrected sequences up to 2 Levenshtein distance away from known messenger sequences.

Fastq read processing for poly-dA oligo capture library

From the fastq file of sequencing the decoder library, the four 10-bp sub-barcodes of the combinatorial spatial barcode were extracted from Read1 and error corrected to known barcodes following the same process as the one for the messenger library. The poly-dA oligo's barcode was extracted from the first ten base pairs of Read2, and the number of UMIs from the Read1 spatial barcode indicated how many poly-dA oligos were captured.

Quantifying bead-bead Interactions

After error-correcting the sequencing reads for the messenger library, we performed 3 steps of filtration to determine real barcodes in our dataset. We first filtered the reads by collapsing interactions between the same pairs of UMIs. Next, we filtered the barcodes by the number of times they appear, keeping barcodes that appear at least "x" number of times, which was determined by where the steepest drop-off occurred in a knee plot of UMI counts. Finally, we filtered for barcodes that appear in both reads sent from beads and reads from receiving beads. We then encoded each bead barcode into a unique integer, and counted the number of interactions between every pair of barcodes. These steps result in an asymmetric interaction count matrix where the rows are barcodes being sent from beads and the columns are barcodes being received by beads. This interaction count matrix was then stored in the sparse matrix format.

Doublet detection to remove sets of beads that share the same DNA barcode

Since large numbers of beads were present in the arrays, a non-negligible number of beads may be doublets, i.e. beads that do not have a unique DNA barcode. The theoretical number of doublets in an array is a function of the number of possible barcodes (N) and the number of beads in the array (n):

$$Expected_number_doublets = \left\lceil \frac{\left(1-\left(1-\frac{1}{N}\right)^{n-1}\right)}{2} \cdot n \right\rceil$$

We performed clustering-based doublet detection to remove such beads. We first built an unweighted graph with the interaction count matrix: if two beads had non-zero interaction counts, there was an unweighted edge between the nodes representing these two beads in the graph. Next, we iterated through each node in the graph and performed Leiden clustering to find communities among its neighbors. The threshold we set for a bead being a doublet was as follows: the neighborhood of the bead should form more than 1 cluster and the largest cluster must be less than 4 times the size of the second largest cluster. Since the number of doublets detected depends on the selected clustering resolution and varies across datasets, for each dataset we performed a grid search on the resolution parameter. We selected the clustering resolution that identified a number of doublets closest to the theoretical value.

Custom simulation of diffusing messenger oligos from beads arrayed in hexagonal lattice

We developed a simulator to guide development of inference methods. To develop the simulator, we did an empirical investigation to understand the decay of bead-bead interaction across distance, and landed on an inverse-square model (**Fig. 2H**). We also used a simple binomial model of bead compatibility. To do a full simulation, we combined the inverse-square weight function with a sample dataframe giving joint row and column read counts to get a per-bead sender and receiver read count, then used multinomial sampling to get the desired number of reads. This ensured that the joint distribution of read counts in the simulation matched those from real data.

Mapping from sender/receiver interactions to pairwise bead distances

Sequencing of the sender-receiver reaction products yields a bead-bead interaction count matrix. This matrix represents the counts of all observed interactions involving an oligo diffusing through space from one bead to another. Thus, it is in essence a “similarity” matrix; higher counts indicate beads that are closer together in space. However, it is not a symmetric measure, as the sending and receiving counts for each pair of beads are represented separately. This similarity matrix is highly sparse, as interactions for most pairs of beads that are not sufficiently close are not observed. Since most methods for reconstructing position data require Euclidean distance matrices as input, we sought to devise a way to map from our similarity matrix, i.e. the measured counts of sender/receiver interactions, to a distance matrix.

Using our custom simulation, we can generate such interaction count matrices alongside simulated bead positions, arranged in a hexagonal lattice. The simulation is parametrized with the per-bead messenger oligo count distribution from the empirical data. After normalizing this count matrix by the total number of molecules sent and received by each bead and averaging the matrix with its transpose, we arrive at a symmetric similarity matrix. We then use a random forest model to perform regression between the normalized pairwise bead-bead similarities for each pair of simulated beads and the corresponding pairwise Euclidean distances from the simulation. We then use this regression function, trained on a simulation, to map our observed, normalized similarity matrix from an experiment to an inferred pairwise distance matrix.

However, this distance matrix has several flaws, chiefly that all the “zeros” in the similarity matrix have been mapped to a single distance value, as per the nature of a regression function. Therefore, our distance matrix is only an approximation of the true pairwise distances between the beads, since the diffusion process that we observe in the form of our interaction matrix only measures discrete molecular binding events. Thus, an important property is that our distance matrix is most accurate at local distances. This precludes the use of classical methods such as Multi-Dimensional Scaling to compute the inferred bead positioning.

Clustering beads and computing cluster-specific reconstructions

Rather, we opted to use t-Distributed Stochastic Neighbor Embedding (t-SNE) to compute the inferred bead positioning. Most often used as a manifold learning algorithm for dimensionality reduction and visualization, t-SNE algorithms can also accept as input a pairwise precomputed distance matrix. Although t-SNE maps are usually said to introduce significant error in dimensionality reduction, our problem is fundamentally a 2-dimension-to-2-dimension mapping problem, and thus avoids the reconstruction error issue. The importance of local distances over global distances, as well as preliminary comparisons on simulations, informed the choice of t-SNE over the Uniform Manifold Approximation Projection (UMAP, **Fig. S3A**).

Given computational restraints as well as the importance of local distances, we chose to approach the problem by subdividing it. Using the normalized similarity matrix as an adjacency matrix, we can construct a graph such that beads are represented by vertices/nodes and edges between them are weighted by the normalized similarity. This graph is then clustered using the Leiden community detection algorithm to group beads into clusters of at most 2500 beads. We then define the relationships between these clusters by defining “boundary beads”- that is, beads who have a certain proportion of edge weights to beads from other clusters. These beads are then defined as belonging to both bordering clusters, creating a set of partially overlapping clusters. These boundary beads are also divided into different sets per cluster, defined by the neighboring cluster that each boundary set borders.

Next, we applied t-SNE to each distance matrix subset from each cluster of beads. However, given that the reconstruction is highly sensitive to the perplexity hyperparameter, we opted for a grid search approach to hyperparameter selection to choose viable solutions. First, each t-SNE reconstruction for each cluster should result in only one “object” and should not form multiple disjoint pieces in space. To assess this, we represent each putative reconstruction (for each hyperparameter value) as a binary image such that each pixel, each spanning an equal area of the reconstruction, is colored black if it has any points within it. This image is then “eroded” using the scikit-image package, followed by image segmentation using the Chan-Vese algorithm. All solutions with more than one object detected are discarded.

We then assess whether the aforementioned “boundary beads” are truly located on the boundary of the inferred reconstruction by first computing the bounding alpha shape (similar to a convex hull) of the reconstruction. This polygon defines the outer boundary of the inferred shape (the outer polygon). Next, we compute the Delaunay triangulation of all points that are not in any boundary bead set and remove layers of edges from the outside of the triangulation to shrink the shape. This shape represents what should be the inner core of the inferred reconstruction (the inner polygon). For each boundary set, we then calculate the proportion of the points that are found within the boundary region between the inner and outer polygons, accepting only the solutions which have over 80% of the points in each boundary set in the geometric boundary region. For example, if the t-SNE reconstruction for a given perplexity value had inverted a portion of the points, the true boundary beads may not actually be found on the inferred geometric boundary (**Fig. S4**).

Finally, we select the final reconstruction for this given cluster from the solutions passing both of these tests by picking the one with the most evenly spaced beads, thus better reflecting the nature of the bead packing on the slide. This is done by computing a 2D histogram of the reconstruction positioning and calculating a chi-squared test statistic comparing this distribution to the uniform distribution. If no perplexity value produced a solution that passed the boundary tests, we subset on the solutions that contain only one "object" and choose the solution with the "least bad" boundary metrics. If all putative solutions contain multiple objects, we do the same procedure and move on, in the interest of eventually arriving at a final global solution. This grid search process is parallelized for each cluster of beads.

Stitching together global reconstruction from individual cluster reconstructions

Once the optimal reconstruction has been calculated for each cluster, we begin the process of "stitching together the puzzle", or assembling the cluster reconstructions in such a way as to reflect the global positioning of all the beads in the same reference frame. Starting from the cluster with the most boundary beads, we stitch one cluster at a time to the growing global reconstruction by learning the parameters of a rigid transformation (rotation, translation, and reflection) by minimizing the mean absolute error between the inferred positions of the boundary beads belonging to both cluster reconstructions through gradient descent. The resulting inferred positioning of all clusters is further refined by learning the parameters of a joint transformation, allowing a separate rigid transformation to be applied to each cluster until all boundary bead positioning errors are minimized. In this way, we arrive at a final set of global positions for each bead. This pipeline thus far represents the "optics-free" mode of spatial genomics.

If more information is provided on bead positions through bead segmentation on brightfield or fluorescence microscopy images, we can then align the inferred positions to the true positions. The alignment can be done by matching the positions of "fiducial" beads, identified through hybridization to an oligo probe complementary to a partial bead barcode sequence. However, since the precise one-to-one matching of fiducial beads is not known, we apply an algorithm for the linear sum assignment problem, or the minimum weight matching problem for bipartite graphs (implemented in scipy). In essence, the set of inferred positions for the fiducial beads can be represented as one partition of vertices, while the true positions represent the other partition. The weights of the edges between these partitions of vertices are defined by the pairwise distances between their respective positions. The rigid transformation that aligns the fiducial bead positions is learned through gradient descent, recomputing the bead point matching at every iteration and minimizing the matching error. After an initial alignment of the fiducials, the point matching is then computed for all points. An optimal rigid transformation to align all the points is then performed and the point matching is recomputed. The final point matching is then used to determine the final positions of the beads. Instead of using the inferred positions of the beads, we use the positions of their inferred match, in effect "snapping" the inferred positions onto the experimental lattice. In this way, we eliminate much of the positioning error as long as the matches are correct. The result from these steps thus represents the more refined result that can be obtained using simple brightfield microscopy and segmentation, and was also applied to results from the simulation in order to quantify reconstruction error (**Fig. S3**).

Downsampling experiment and evaluation of t-SNE vs UMAP

In order to determine how the reconstruction accuracy varies with the sequencing depth of the sender/receiver library, we designed a synthetic downsampling experiment. We simulated a 40,000 bead array using the framework above, parametrized using the counts of the asymmetric array experiment. We then downsampled the total counts of the simulated bead-bead interaction matrix to various levels ranging from 1% to 100% of the

total counts. For each chosen downsampling percentage, we attempted to reconstruct the spatial positions of the beads and align them to the true known positions of the beads according to the simulation, following the process as we would in a true experiment with mapped fiducials and unknown bead barcode positions. We repeated this process 10 times for each downsampling percentage and computed the mean absolute error of the inferred bead positions after alignment (**Fig. S3**). On the x-axis, we plot the mean messenger UMI counts per bead after downsampling.

In order to evaluate the accuracy of t-SNE vs UMAP at various scales, we simulated rectangular hexagonal lattice bead arrays at various different sizes and computed the t-SNE and UMAP reconstructions on this bead-bead interaction data without breaking up the graphs into clusters. The reconstructions were then registered and point matched to the ground truth as above, and the mean absolute error was calculated. The median and interquartile range of 5 trials were plotted for each simulated array size.

Reconstructing the letters in the Snellen eye chart

For the Snellen diagram experiment, the global reconstruction was not optimal due to error propagation issues in the cluster stitching step (**Fig. S8**). In order to determine if certain regions of the solution were faithfully reconstructed, we first plotted the counts of captured printed oligos to see if the letters were discernible. Since the large “E” was quite visible, we then subsetted the beads in that region based on the coordinates of the initial global reconstruction and then re-computed a new reconstruction on just those beads, only considering sender/receiver information among those pairs. This reconstruction of just the cropped big “E” region was run several times and the best result was chosen for visualization (**Fig. 4B**). This process was repeated for the other letters. The binned image was computed by tiling the reconstruction point cloud with equally sized squares and summing up the squared counts of the captured DAPI oligo (**Fig. 4D**). Horizontal and vertical line plots were computed using the values in the bins that the lines intersected at each axis coordinate.

Acknowledgments
This work was supported by a grant from the Paul G. Allen Frontiers Group (Allen Discovery Center for Cell Lineage Tracing, to J.S.), Alex's Lemonade Stand Foundation (to J.S.), the National Institutes of Health (HG010632 to J.S. and C.T., AI146028 to F.A.M.), as well as philanthropic support from the Brotman Baty Institute for Precision Medicine. H.L. is supported by the NSF Graduate Research Fellowship Program (DGE-2140004). J.S. and F.A.M. are Investigators of the Howard Hughes Medical Institute.

Data and Code Availability
Downsampled sequencing data for the reconstruction experiments have been deposited to Github (https://github.com/SrivatsanLab/SCOPE), while the raw data files are available via Dropbox. Our custom reconstruction pipeline used to reconstruct the SCOPE experiments is also available at Github (https://github.com/SrivatsanLab/SCOPE), and simulation code (https://github.com/matsengrp/sci-space-v2)

Competing interests
The Fred Hutchinson Cancer Center and University of Washington have filed a patent application partially based on this work, on which S.S., J.S., H.L., J.G. and S.K. are listed as inventors. J.S. is on the scientific advisory board, a consultant, and/or a co-founder of Prime Medicine, Cajal Neuroscience, Guardant Health, Maze Therapeutics, Camp4 Therapeutics, Phase Genomics, Adaptive Biotechnologies, Scale Biosciences, Sixth Street Capital, Pacific Biosciences and Somite Therapeutics. The other authors have no competing interests to declare.

References

1. C. G. Williams, H. J. Lee, T. Asatsuma, R. Vento-Tormo, A. Haque, An introduction to spatial transcriptomics for biomedical research. *Genome Med.* **14**, 68 (2022).

2. L. Moses, L. Pachter, Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).

3. K. Vandereyken, A. Sifrim, B. Thienpont, T. Voet, Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023).

4. A. J. C. Russell, J. A. Weir, N. M. Nadaf, M. Shabet, V. Kumar, S. Kambhampati, R. Raichur, G. J. Marrero, S. Liu, K. S. Balderrama, C. R. Vanderburg, V. Shanmugam, L. Tian, J. B. Iorgulescu, C. H. Yoon, C. J. Wu, E. Z. Macosko, F. Chen, Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature* **625**, 101–109 (2024).

5. S. G. Rodriques, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, E. Z. Macosko, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

6. S. R. Srivatsan, M. C. Regier, E. Barkan, J. M. Franks, J. S. Packer, P. Grosjean, M. Duran, S. Saxton, J. J. Ladd, M. Spielmann, C. Lois, P. D. Lampe, J. Shendure, K. R. Stevens, C. Trapnell, Embryo-scale, single-cell spatial transcriptomics. *Science* **373**, 111–117 (2021).

7. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

8. J. A. Weinstein, A. Regev, F. Zhang, DNA Microscopy: Optics-free Spatio-genetic Imaging by a Stand-Alone Chemical Reaction. *Cell* **178**, 229–241.e16 (2019).

9. J. Y. Kishi, T. E. Schaus, N. Gopalkrishnan, F. Xuan, P. Yin, Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem.* **10**, 155–164 (2018).

10. J. Davis, Microvenus. *Art Journal* **55**, 70–74 (1996).

11. G. M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012).

12. J. Choi, W. Chen, A. Minkina, F. M. Chardon, C. C. Suiter, S. G. Regalado, S. Domcke, N. Hamazaki, C. Lee, B. Martin, R. M. Daza, J. Shendure, A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98–107 (2022).

13. W. Chen, J. Choi, X. Li, J. F. Nathans, B. Martin, W. Yang, N. Hamazaki, C. Qiu, J.-B. Lalanne, S. Regalado, H. Kim, V. Agarwal, E. Nichols, A. Leith, C. Lee, J. Shendure, Symbolic recording of signalling and cis-regulatory element activity to DNA. *Nature*, doi: 10.1038/s41586-024-07706-4 (2024).

14. G. M. Church, S. Kieffer-Higgins, Multiplex DNA sequencing. *Science* **240**, 185–188 (1988).

15. P. W. K. Rothemund, Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).

16. Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, E. Shapiro, Programmable and autonomous computing machine made of biomolecules. *Nature* **414**, 430–434 (2001).

17. J. Y. Kishi, N. Liu, E. R. West, K. Sheng, J. J. Jordanides, M. Serrata, C. L. Cepko, S. K. Saka, P. Yin,

Light-Seq: light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing. *Nat. Methods* **19**, 1393–1402 (2022).

18. A. Chen, S. Liao, M. Cheng, K. Ma, L. Wu, Y. Lai, X. Qiu, J. Yang, J. Xu, S. Hao, X. Wang, H. Lu, X. Chen, X. Liu, X. Huang, Z. Li, Y. Hong, Y. Jiang, J. Peng, S. Liu, M. Shen, C. Liu, Q. Li, Y. Yuan, X. Wei, H. Zheng, W. Feng, Z. Wang, Y. Liu, Z. Wang, Y. Yang, H. Xiang, L. Han, B. Qin, P. Guo, G. Lai, P. Muñoz-Cánoves, P. H. Maxwell, J. P. Thiery, Q.-F. Wu, F. Zhao, B. Chen, M. Li, X. Dai, S. Wang, H. Kuang, J. Hui, L. Wang, J.-F. Fei, O. Wang, X. Wei, H. Lu, B. Wang, S. Liu, Y. Gu, M. Ni, W. Zhang, F. Mu, Y. Yin, H. Yang, M. Lisby, R. J. Cornall, J. Mulder, M. Uhlén, M. A. Esteban, Y. Li, L. Liu, X. Xu, J. Wang, Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).

19. X. Fu, L. Sun, R. Dong, J. Y. Chen, R. Silakit, L. F. Condon, Y. Lin, S. Lin, R. D. Palmiter, L. Gu, Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* **185**, 4621–4633.e17 (2022).

20. I. T. Hoffecker, Y. Yang, G. Bernardinelli, P. Orponen, B. Högberg, A computational framework for DNA sequencing microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19282–19287 (2019).

21. A. Kloosterman, I. Baars, B. Högberg, An error correction strategy for image reconstruction by DNA sequencing microscopy. *Nat Comput Sci* **4**, 119–127 (2024).

22. D. Fernandez Bonet, I. T. Hoffecker, Image recovery from unknown network mechanisms for DNA sequencing-based microscopy. *Nanoscale* **15**, 8153–8157 (2023).

23. F. Karlsson, T. Kallas, D. Thiagarajan, M. Karlsson, M. Schweitzer, J. F. Navarro, L. Leijonancker, S. Geny, E. Pettersson, J. Rhomberg-Kauert, L. Larsson, H. van Ooijen, S. Petkov, M. González-Granillo, J. Bunz, J. Dahlberg, M. Simonetti, P. Sathe, P. Brodin, A. M. Barrio, S. Fredriksson, Molecular pixelation: spatial proteomics of single cells by sequencing. *Nat. Methods* **21**, 1044–1052 (2024).

24. N. Qian, J. A. Weinstein, Volumetric imaging of an intact organism by a distributed molecular network. *bioRxiv*, doi: 10.1101/2023.08.11.553025 (2023).

25. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction (2018). http://arxiv.org/abs/1802.03426.

26. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

27. C. L. Delley, A. R. Abate, Modular barcode beads for microfluidic single cell genomics. *Sci. Rep.* **11**, 10857 (2021).

28. M. E. Fornace, J. Huang, C. T. Newman, N. J. Porubsky, M. B. Pierce, N. A. Pierce, NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems. *ChemRxiv*, doi: 10.26434/chemrxiv-2022-xv98l (2022).

29. D. A. Cusanovich, R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell, J. Shendure, Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).

30. C. G. Evans, J. O'Brien, E. Winfree, A. Murugan, Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly. *Nature* **625**, 500–507 (2024).

31. M. Zhang, X. Pan, W. Jung, A. R. Halpern, S. W. Eichhorn, Z. Lei, L. Cohen, K. A. Smith, B. Tasic, Z. Yao, H. Zeng, X. Zhuang, Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).

32. J. Langlieb, N. S. Sachdev, K. S. Balderrama, N. M. Nadaf, M. Raj, E. Murray, J. T. Webber, C. Vanderburg, V. Gazestani, D. Tward, C. Mezias, X. Li, K. Flowers, D. M. Cable, T. Norton, P. Mitra, F. Chen, E. Z. Macosko, The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333–342 (2023).

33. F. A. Wolf, F. K. Hamey, M. Plass, J. Solana, J. S. Dahlin, B. Göttgens, N. Rajewsky, L. Simon, F. J. Theis, PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).

34. R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, L. Mazutis, Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).