From whole-mount to single-cell spatial assessment of gene expression in 3D

Lisa N. Waylen1, Hieu T. Nim1,2, Luciano G. Martelotto3 & Mirana Ramialison1,2✉

Unravelling spatio-temporal patterns of gene expression is crucial to understanding core biological principles from embryogenesis to disease. Here we review emerging technologies, providing automated, high-throughput, spatially resolved quantitative gene expression data. Novel techniques expand on current benchmark protocols, expediting their incorporation into ongoing research. These approaches digitally reconstruct patterns of embryonic expression in three dimensions, and have successfully identified novel domains of expression, cell types, and tissue features. Such technologies pave the way for unbiased and exhaustive recapitulation of gene expression levels in spatial and quantitative terms, promoting understanding of the molecular origin of developmental defects, and improving medical diagnostics.

Spatial control of gene expression is crucial for defining tissue identity, from the patterning of a zebra’s stripes, to the left-right asymmetry present in most species, and deviations from this regulatory programme may result in developmental defects or disease, such as situs inversus or the variation in antigens expressed at different layers of a cancer tumour. Our body plan relies on spatial expression, achieved by correct deployment of a developmental gene regulatory network (GRN) where the location, timing, and level of developmental gene expression are crucial. For instance, Hox genes are a highly conserved family involved in body axis patterning and provide an illustration of the requirement for specific timing, spatial restriction, and expression degree across species. Hox genes are silenced until activated by developmental timing cues, and sequential expression of Hox clusters will define the body plan of the embryo1–3. Ectopic expression of Hoxa-1 in mice can reorganise regions of the hindbrain by transforming developing rhombomeres4, and in zebrafish, misexpression of Gata5 induces ectopic expression of myocardial genes such as nkx2.55. Of particular importance in development are morphogens, a concentration gradient dependent signal that provides a pattern for tissue differentiation and development. While some genes employ a simple on/off mechanism, many developmental genes respond to different expression thresholds to provide phenotypically distinct outcomes. In Drosophila, the maternal gene bicoid is responsible for regulating
development of the anterior pole through graduated mRNA diffusion\(^6\) and limb patterning in the chick is established by morphogenetic fields\(^7\).

Organoids raised from human stem cell cultures provide an invaluable model for exploring disease mechanisms spatially, by uncovering the cellular and molecular interactions occurring in three dimensions (3D). For instance, embryonic models such as gastruloids\(^8\), and subdomains of organs such as the kidney can be recapitulated from the preferential induction of progenitors\(^9\).

This offers a unique platform to capture the spatial gene expression information required for accurate tissue patterning. Deviations from programmed, spatio-temporal gene expression can lead to disease, as seen in congenital disorders\(^10\) and highly heterogeneous cancer tumours which provide a prime target for spatial profiling analysis. For example, the dysregulation of HOX genes in leukaemia has been shown to support the immortalisation of malignant cells, while specific spatial expression patterns of biomarkers inform clinical prognosis and therapies\(^11,12\).

In plants, changes in spatial gene expression have been shown to provide advance warning for the progression of viral infection\(^13\). Therefore, to accurately decipher the GRNs involved in spatial gene expression, it is necessary to capture gene identity along with quantitative data. Previously, the Allen Mouse Brain Atlas constructed by the Allen Institute for Brain Sciences\(^14\) integrated bulk and computational methods with in situ hybridisation gene expression data, later accompanied by human and macaque atlases\(^15,16\). Despite challenging low resolution and lack of quantitative data, these atlases significantly advanced the current understanding of brain structure and function. More recently, the Allen Mouse Brain Common Coordinate Framework (CCFv3)\(^17\) sampled 1675 mouse brains at 10 \(\mu m\) voxel resolution, integrating multiple datasets in 3D. Virtual atlases are publicly available, browsable cell reference maps where gene expression data is reconstructed in 3D, digital space. They support myriad fields of research through interactive visualisation and analysis, providing a valuable source of information for many model organisms including\(C.\ elegans\) (Wormbase)\(^18\), frogs (Xenbase)\(^19\), Drosophila\(^20\) and zebrafish\(^21\).

This review explores current techniques used to elucidate spatial expression, from established robust methods with throughput limitations, to cutting-edge systematic and unbiased measurement of gene expression (Fig. 1, Table 1). We explain the power of accurately capturing spatial gene expression data quantitatively, to decipher molecular interactions driving major biological processes.

### Imaging-based methods for resolving gene expression spatially

Historically, spatial patterns were captured from direct or in situ staining of genes of interest. In situ hybridisation (ISH) techniques (Fig. 1a) were developed to study spatial complexity. This involves hybridisation of a labelled RNA to an endogenous mRNA transcript within the cell\(^22\). Large-scale in situ hybridisation is routinely used to simultaneously interrogate expression patterns of genes in sections or whole organs and tissues\(^23,24\).

This is a particularly efficient method to screen for genes with regionalised expression patterns, such as those responsible for early growth determination\(^25\), but these techniques are low-throughput. At most a hundred genes can be simultaneously assessed, albeit in different samples. Another challenge of this traditional model is that visualisation of expression relies on probe binding efficiency, which varies significantly, eliminating the ability to perform comparative study between tissues or genes, without first performing background neutralising or equalising equations\(^26\). While in situ hybridisation addresses the need for spatially resolved data, accurate quantification of gene expression is often difficult to achieve using this approach and therefore requires further analysis.

### Increasing sensitivity and accuracy

Further advancements to accurate quantitative measurement of in situ gene expression include the Spatial Genomic Analysis pipeline (SGA)\(^27\), which integrates the output from sequential single-molecule fluorescence in situ hybridisation and hybridisation chain reaction (HCR). HCR employs two complementary DNA probes which form a fluorescently labelled amplification polymer capable of relative quantification of mRNA expression\(^28\). In addition, this technology achieves higher spatial resolution due to the integrated single-cell imaging techniques. The SGA pipeline allowed novel characterisation of a pluripotent stem cell niche in development of the dorsal neural tube\(^29\). RNAscope provides a further integration of qualitative and quantitative approaches whereby probe hybridisation and signal amplification are performed in situ, allowing spatially relevant analysis at single-molecule resolution\(^30\). However, analysis is limited to a handful of target genes, since throughput from SGA is restricted by the number of distinguishable dyes.

Click-amplifying fluorescent ISH (ClampFISH) was developed to improve the efficiency in FISH and the resulting weak signal response that demands high power microscopy. Indeed, the enzymes required in FISH may prove inefficient, providing inconsistent amplification or poor diffusion through the cell. ClampFISH detects unique nucleic acid molecules, and allows padlock-style probes to be used without enzyme support. Click chemistry is used to covalently circularise amplifier molecules by linking 5' and 3' ends\(^30\), topologically entangling individual amplifier molecules to increase binding specificity.

Recent advances in ISH have introduced a novel approach to improve signal detection. Single-molecule fluorescence in situ hybridisation (smFISH) relies on the use of many short probes which target multiple regions of the mRNA transcript. This can approximate the level of gene expression between different tissues, but is not numerically quantitative\(^31\). It is also limited by the need to bind sufficient primers to the query mRNA, which may not be possible for many short mRNAs. Although smFISH cannot assess non-coding RNAs such as those required for many regulatory functions, as it targets the ribosome/mRNA interactions, it can provide spatial localisation of RNA expression by directly imaging individual RNA molecules in single cells. Throughput is limited by the number of RNA species that can be simultaneously measured in single cells.

While previous studies in smFISH have improved the number of probed genes accessible via a targeted, barcode approach, a new cyclic technique, ouroboros smFISH (osmFISH) relies on a non-barcoded approach focussed on designing image processing tools capable of processing large tissue areas and extensive data sets\(^32\). osmFISH is more efficient at detecting RNA than single-cell RNA sequencing (scRNA-seq)\(^33\), and osmFISH has combined short hybridisation times and background reduction to attain high signal-to-noise ratios to further improve the sensitivity of RNA detection. osmFISH achieved a rate of 30% zero counts in contrast to the 82% registered by scRNA-seq, indicating its higher sensitivity to recover low levels of gene expression. This is largely because target number scales with the number of required hybridisation rounds to reduce error, and highly expressed genes do not affect the detection of non-highly expressed genes as barcoding is not required, and images can be analysed separately. FISH approaches rely on downstream image analysis to extract numerical measurements of gene expression. Meanwhile, scRNA-seq also requires reads filtering, mapping, and quality control to
be converted into numerical measurements. *osmFISH* has been optimised for thin tissue sections, and a study of murine brain tissue developed an atlas of gene expression in the somatosensory cortex using 33 marker genes. Although a majority (59%) of the data came from neurons, many different cell types were observed and characterised.

**Increasing throughput and resolution.** Multiplexed error-robust fluorescence in situ hybridisation (MERFISH) multiplexes *smFISH* and assigns barcodes to individual RNA species, which can be imaged sequentially, allowing single-cell transcriptomic profiling. Branched DNA (bDNA) amplification substantially improved MERFISH detection efficiency, without increasing fluorescent spot size for individual transcripts, and it is now possible to use this method to study short RNAs. This approach generates in situ transcriptomic analyses and detects >1000 RNA species, and provides error correction for >100. Further improvements to the efficiency of MERFISH gene processing allow simultaneous imaging of ~10,000 genes at a detection efficiency of ~80%. While providing a high standard of subcellular resolution, the necessity for cell-dissociation challenges the spatial power of this approach.

Digital Spatial Profiling (Fig. 1b), a new platform from Nanostring, allows multiplexed analysis of protein or RNA (~100 and ~1000plex, respectively), capable of single-cell resolution. Probes coupled with photocleavable, spatially barcoded tags can be used to simultaneously profile targets from a sample tissue region. Early testing suggests that this will be a powerful new tool in understanding the tumour microenvironment and identifying unique, prognostic patient biomarkers. Reproducibility between sections is high, and result validation demonstrates a fair correlation with IHC, FCM and QIF, with some spatially restricted expression patterns of immune markers found to be associated with patient outcome. A key advantage of this approach is that it is non-destructive, permitting reanalysis, and analysis of protein and RNA in cohort on serial sample sections. Digital Spatial Profiling is currently limited, with only four fluorescent visualisation stains available to examine tissue morphology and determine regions of interest.

Traditional spatial mapping relies on fluorescence light microscopy or direct physical registration of transcripts. DNA
microscopy (Fig. 1c) instead, involves chemical reactions whereby transcript molecules are tagged in situ with random nucleotides, labelling each uniquely39. A second reaction amplifies these tagged molecules, concatenates the copies, and adds new random nucleotides. While high signal intensity is a drawback to fluorescent imaging techniques, increased signal density is beneficial to DNA microscopy. Molecular proximities are computationally decoded from the overlapping diffusion fields generated by these signals, from which it is possible to infer physical images of the original transcripts at cellular resolution. Sequencing power permits detection of single-nucleotide variation, and spatially resolves biological features such as somatic mutation and stochastic RNA splicing. Variation in molecular density, however, makes it difficult to reconstruct images spanning large distances. Early applications indicate that DNA microscopy can function as an imaging medium equivalent to optical analysis, and is an important emerging approach to study the complex chemical dynamics of biological systems.

FISH-based approaches allow transcripts to be directly labelled within native tissue sections. In regions of dense transcript concentration, molecular crowding leads to the spatial overlap of fluorescence signals during simultaneous transcript imaging. Extended sequential FISH (seqFISH+) serially reprograms a single slide to generate multiple images which can be computationally merged and analysed for a complete library of transcripts (Fig. 1d), reducing the impact of optical crowding as only 1/100th of the transcript is visualised in each image40. In one study, 10,000 genes were analysed in cultured mouse cells. The subsequent gene expression profiles were quantitatively equivalent to RNA-seq data. Within the brain, cells were characterised based on their expression profile, revealing the expected tissue layers when analysing the spatial relationships. Ligand-receptor encoding RNA pairs observed in adjacent cells indicated functional relationships which were dependent on local tissue context/positioning. Comparison with the 60 genes analysed from Drosophila eye, including interommatidial cells and photoreceptors, each exhibiting a unique set of differentially expressed genes. This map provides a transcriptome-wide reflection of spatial patterning which was used to identify key cells subpopulations in the Drosophila eye, including interommatidial cells and photoreceptors, each exhibiting a unique set of differentially expressed genes.

In summary, in situ mapping of spatial gene expression captures both RNA quantity and position. It has overcome challenges to maintain efficiency, signal intensity and accuracy while scaling to large gene numbers. However, these approaches are still limited in their throughput and ability to accurately measure the level of gene expression.

Table 1 Features of current methods for capturing spatial gene expression.

| Methodology                | Coverage       | Number of genes | Number of cellsa | Spatial resolution |
|----------------------------|----------------|-----------------|------------------|--------------------|
| In situ hybridisation32    | Targeted       | 3               | Low              | Tissue             |
| RNAscope29                 | Targeted       | 12              | Low              | Cellular           |
| ClampFish30                | Targeted       | 3               | Low              | Subcellular        |
| smFISH33                   | Targeted       | 3               | Low              | Subcellular        |
| osmFISH32                  | Targeted       | 1-33+           | Low              | Subcellular        |
| MERFISH55                  | Targeted       | 10,000          | Medium           | Subcellular        |
| DNA microscopy39           | Targeted       | 3               | Low              | Cellular           |
| seqFISH+40                 | Targeted       | 10,000          | High             | Subcellular        |
| DistMap50                  | Targeted       | 8000+           | High             | Cellular           |
| STARMap51                  | Targeted       | 1020+           | High             | Cellular           |
| Tomo-seq52                 | Transcriptome-wide | Whole transcriptome | High         | Cellular           |
| Geo-seq56                  | Transcriptome-wide | Whole transcriptome | High         | Cellular           |
| Spatial transcriptomics/10X Visium59 | Transcriptome-wide | Whole transcriptome | Medium       | 100 µm/55 µm      |
| Slide-seq61                | Transcriptome-wide | Whole transcriptome | Medium       | 10 µm              |
| HDST63                     | Transcriptome-wide | Whole transcriptome | High          | 2 µm               |
| novoSpaRc59                | Transcriptome-wide | Whole transcriptome | High          | Cellular           |
| NASC-seq52                 | Transcriptome-wide | Whole transcriptome | High          | Cellular           |

aNumber of cells—low: 0–100, medium: 100–1000, high: 1000–10,000–.

Next-generation sequencing-based methods for resolving gene expression spatially

Quantitative methods of measuring accurate gene expression levels in high-throughput have been developed such as CAGE, SAGE, and gene expression arrays. CAGE sequencing permits analysis of coding and non-coding RNA expression by targeting promoter regions instead of translation interactions41. Massive parallel RNA sequencing (RNA-seq) allows a high-throughput accurate measurement of RNA transcript levels that can be achieved in an organ (bulk)42 or single cells through read-counts43. Data from bulk tissue analysis may mask the expression patterns of rare, or equally dominant cell types, or omit transient expression patterns. The quantitative data gained in RNA-seq is contrasted by the loss of spatial information. When isolated cells are spatially dislocated from their tissue of origin, transcripts lose spatial context; however, some studies have attempted to address these issues by using downstream cell sorting which reunites distinct cell populations. To achieve an analysis which benefits from data obtained from both spatial and quantitative methods, the two data sets need to be combined, a process which remains largely manual. Recent experiments to combine spatial information, with the quantitative data obtained from RNA-seq have been used to query the transcriptional changes in different regions of the cardiac infarct zone following heart injury in adults and neonates, providing type-specific quantitative expression data44.

Automated hybrid approaches. Mathematical models allow the integration of spatial and quantitative results by associating ‘landmark’ reference genes with transcript counts. For instance, Seurat is an algorithm that infers cellular localisation by integrating scRNA-seq data with in situ expression patterns45. It relies on spatial segregation of landmark genes, identified through ISH experiments, to construct a reference map by including the RNA-seq data to quantify the expression level of these landmark genes. This map provides a transcriptome-wide reflection of spatial patterning which was used to identify key cells subpopulations in the Drosophila eye, including interommatidial cells and photoreceptors, each exhibiting a unique set of differentially expressed genes.

Moreover, one of the primary problems associated with scRNA-seq analysis, is the high technical noise of small-mass embryos, and subsequent low-copy number transcripts. To
overcome this, the computations in Seurat also consider the spatial information of multiple genes that are co-regulated with landmark genes, thus providing a reliable gene expression map. Another computational method which integrates ISH with scRNA-seq provides a high-throughput approach to identify the tissue of origin for complex tissues.

High-throughput tissue sectioning-based approaches. Tomo-seq (Fig. 1g) automatically generates spatial resolution equivalent to in situ hybridisation with the quantitative strength of RNA-seq. The tissue of interest is cryosectioned in three different orientations (antero-posterior, dorso-ventral and lateral planes), followed by bulk RNA-seq of each section. Triangulation algorithm, RNA-tomography, is then applied to estimate the gene expression levels at the intersection of the three sections, creating a digital three-dimensional expression pattern of the whole embryo. Depending on section thickness, this mathematical reconstruction could recapitulate gene expression at a resolution equivalent to single-cell analysis. However, the accuracy of this computed gene expression value needs to be validated with scRNA-seq datasets. Tomo-seq necessitates using identical samples for an accurate overlay of sections on multiple axes. This is biologically difficult to accomplish, there is room for a degree of error in combining tissue sections. Applied to developing zebrafish embryos at three developmental stages (shield, 10 somites and 15 somites stages), Tomo-seq delivers a comprehensive digital resource for embryonic gene expression in the whole embryo. Approximately 20% of the zebrafish transcriptome exhibited spatially restricted gene expression, a large number of which were functionally uncharacterised.

Combined with chromosome conformation capture datasets, Tomo-seq was used to identify spatially co-expressed genes correlating with topological domains. Pre-zygotic transcription exhibited strict structuring with super enhancers clustering during development, a pattern which is similar to that previously observed in mammals. Super enhancer regions present a greater degree of transcriptional regulation, and are associated with more highly expressed genes. Tomo-seq has also been successfully applied to an entire isolated organ. Indeed, a high spatial resolution map of the embryonic zebrafish heart obtained with Tomo-seq revealed 1100 genes being differentially expressed in the various sub-compartment of the developing heart, 502 of which were previously uncharacterised. Further examination of the sinoatrial pacemaker region of the heart revealed a novel role for Wnt/β-catenin signalling in the pathway regulating autonomic control, which provided useful insights into the molecular basis of heart function. This illustrates the discovery potential of this technique to search for novel gene targets for therapeutic use.

A complementary method, Geo-seq (Fig. 1h), cryosections organs or tissues of interest in one direction only. Each section is sampled by laser capture microdissection, generating cell clusters (~10 cells per cluster), which are then subjected to bulk RNA-seq. As the spatial origin of each cell cluster is uniquely identified, this allows reconstruction of spatial distribution of gene expression within the whole tissue using the Zipcode Mapping Feature. Geo-seq was applied to gastrulating mouse embryos to generate a virtual whole-mount 3D model of the spatial expression of all 20,000 mouse genes at this stage. This iTranscriptome was central in revealing novel molecular markers for different cell lineages, for instance, highlighting the role of Oct4 in cell fate determination and A–P axis patterning.
Spatial transcriptomics (Fig. 1i) is the least biased method to capture spatial gene expression from sections. cDNA is directly synthesised from fixed tissue sections on 6.2 mm × 6.6 mm pucks and is concomitantly labelled with molecular barcodes which record the spatial location of the transcript within the section59. The cDNA is then sequenced by RNA-seq. This approach eliminates the need for cell isolation, and was able to detect low-level transcript expression and capture more genes and transcripts than laser capture microdissection, and with higher sensitivity and resolution compared to Tomo-seq and Geo-seq. The ability to generate histological data which precisely complements the tissue section involved in quantitative analysis is unique to Spatial transcriptomics (microarray-based technology) and not possible with NGS-based methods. By processing different tissue domains within the same reaction, Spatial transcriptomics is able to remove much of the technical variation experienced by many standard expression methods. Application of this technique in murine brain tissue demonstrated gene clustering that correlated with morphological layers, allowing identification of cluster (and subsequently tissue) specific markers, while analysis of human breast cancer cells revealed that invasive cells exhibited high expression of extracellular-associated genes, and an unexpectedly high degree of cellular heterogeneity within biopsy samples, information important for decisions about prognosis59. The integration of Spatial transcriptomics with scRNA-seq from dissociated whole human embryonic hearts was able to identify a novel population of MyoZ2-enriched cardiomyocytes in both the atria and ventricles which complemented recent findings in mice, yet to be observed in humans60. Currently, the resolved power of Spatial transcriptomics remains limited by the spacing of primers required for barcoding.

Slide-seq (Fig. 1j) alleviates this limited resolution using DNA-barcoding beads bound to 3-mm slides and exposed to fresh tissue sections releasing mRNA from which barcode sequences can be determined with sequencing by oligonucleotide ligation and detection (SOLID)61. This variant of Drop-seq62 allowed the identification of fine spatial features including single-cell layers in the mouse brain61. It revealed cell zonation patterns and cellular constituents, consistent with findings observed in human post mortem brain tissue. Tissue structure dimensions did not change, indicating minimal lateral diffusion from the barcoded beads. Analysis of gene expression following traumatic brain injury revealed cell proliferation progressing to differentiation. This technique accurately reflected spatial distribution of classical neuronal and non-neuronal cell types, ~66% of the DNA-barcoding beads matched to single-cell types, while ~33% matched to two cell types. Expression values concurred with results from bulk-mRNA-seq and scRNA-seq, and average mRNA capture was consistent across all tissues61. Advancements have further improved spatial resolution, such as high-definition spatial transcriptomics (HDST) which employs 2 μm spatial barcoding63 and permits histological analysis which is not possible with Slide-seq.

Computational approaches for resolving spatial gene expression

Integrating spatial and expression information. It is possible to map scRNA-seq data onto Slide-seq data using non-negative matrix factorisation regression (NMReg) which reconstructs Slide-seq expression as a combination of cell-type signatures from scRNA-seq64. LIGER (linked inference of genomic experimental relationships) is a computational method for spatially locating cells present in scRNA-seq data from in situ transcriptomic data, thereby increasing the resolution of the in situ data64. The Harmony algorithm, which projects cells into a shared embedding in which cells group by cell type rather than dataset-specific conditions, is shown to be both efficient and accurate65. Another multiple datasets integration pipeline, inspired by multiple sequence alignment, utilises canonical correlation analysis (CCA) to identify anchor points across heterogeneous datasets66. An alternative approach is SpaOTsc, which utilises genes with spatial measurements to extrapolate the spatial properties of scRNA-seq data67. A different approach is found in Giotto, a user-friendly workspace that utilises cell-type-specific gene signatures to infer cell-type enrichment scores for downstream analyses with the capability for integration of spatial information68. Altogether, these computational approaches present timely developments to capitalise on the rapid growth of high-throughput tissue sectioning-based technologies and are summarised in Table 2. However, these methods cannot yet incorporate multiple data types, such as gene expression and intergenic methylation, in defining cell types.

De novo spatial position prediction using only expression data. Rather than combining two different datasets, a new class of advanced computational techniques enables the prediction of spatial information from a single gene expression dataset. novoSpaRc is a recent gene expression cartography technique that performs mapping based on the variation of gene expression across a tissue section, using a probabilistic optimisation technique69. However, the accuracy can be sub-par due to the lack of reference map, and a set of a priori marker genes with known expression patterns is desirable. In DrosoPhila, ScoMAP (Single-Cell Omics Mapping into spatial Axes using Pseudotime ordering) is another reference-free technique that spatially integrates expression data into a virtual latent space, resembling the organization of a 2D tissue70. At the single-cell resolution level, the CSOmap (Cellular Spatial Organization mapper) algorithm can partially reconstruct the tissue spatial organisation based on ligand-receptor interaction71. Altogether, these computational advances demonstrate the potential power of in silico techniques in de novo spatial mapping of expression data, but there is still room for improvement in prediction accuracy.

High-throughput native single-cell approaches. Recently, the emergence of high-throughput single-cell resolution 3D analysis methods has provided the spatio-temporal expression information necessary to understand embryogenesis and development. The main advantage of these techniques over the previous integrating methods, is that these do not require prior annotation, nor do they rely on tissue dissociation to obtain quantitative values.

novoSpaRc (Fig. 1k) is a new computational framework which allows de novo spatial reconstruction of single-cell gene expression69. Previous computational approaches have required a reference atlas of marker genes to assign spatial coordinates. Prior knowledge is not required for novoSpaRc, as this method takes single-cell transcriptome profiles, sequenced from dissociated cells as its input, then returns a virtual tissue of a chosen shape, which can be queried for the expression of all genes quantified in the data. There is a high degree of variability in how these sequenced cells may be arranged. The novoSpaRc hypothesis is that genes are often expressed in spatially contiguous territories, and by tracking cells exhibiting similar features, the best matching spatial arrangement of cells can be found. Spatial patterns are reconstructed with little prior knowledge, by looking for the spatial arrangement of sequenced cells in which nearby cells have transcriptional profiles that are often more similar than cells that are further apart.
| Tool           | Underlying method                                                                 | Open-source | Input                                                                 | Output                                                                 | Programming language | Source code                                      |
|---------------|-----------------------------------------------------------------------------------|-------------|-----------------------------------------------------------------------|------------------------------------------------------------------------|----------------------|-------------------------------------------------|
| RNAscope⁹⁹    | Proprietary software                                                              | No          | N/A                                                                   | N/A                                                                   | N/A                  | N/A                                             |
| DistMap⁵⁰     | Distributed mapping scores                                                        | Yes         | Count matrix, reference in situ coordinates                           | Expression patterns                                                    | R                    | https://github.com/rajewsky-lab/distmap         |
| NMFreg⁶¹      | Non-negative matrix factorization regression                                        | Yes         | Count matrix, 2D spatial coordinates                                   | Expression patterns                                                    | Python               | https://github.com/tudaga/NMFreg_tutorial       |
| LIGER⁶⁴       | Integrative non-negative matrix factorization                                     | Yes         | Count matrix, scMethylation, scATAC-seq                                | Expression patterns                                                    | R                    | https://github.com/MacoskoLab/liger            |
| Harmony⁶⁵     | Maximum diversity clustering, linear batch correction                              | Yes         | Count matrix, MERFISH coordinates                                      | Expression patterns                                                    | R                    | https://github.com/immunogenomics/harmony       |
| SpotOTsc⁶⁷    | Structured optimal transport                                                       | Yes         | Count matrix, spatial coordinates, dissimilarity matrices             | Expression patterns                                                    | Python               | https://github.com/zcang/SpaOTsc                |
| NovoSpaRc⁶⁹   | Generalised optimal-transport                                                       | Yes         | Count matrix, target space image                                       | Expression patterns                                                    | Python               | https://github.com/rajewsky-lab/novosparc       |
| ScoMAP⁷⁰      | Axial information extraction via pseudotime ordering                               | Yes         | Count matrix, virtual spatial template                                 | Expression patterns                                                    | R                    | https://github.com/aertslab/ScoMAP             |
| CSOmap⁷¹      | Cellular Spatial Organization mapper                                               | Yes         | Count matrix, label, ligand-receptor                                   | Expression patterns                                                    | MatLab               | https://codeocean.com/capsule/2860903/tree/v1  |
| Giotto⁶⁸      | Binary Spatial extraction, hidden Markov random field (HMRF) model                 | Yes         | Count matrix, spatial coordinates                                      | Expression patterns                                                    | R, Python            | https://rubd.github.io/Giotto/                 |
| SPOTlight⁹⁸   | Seeded non-negative matrix factorization (NMF) regression                           | Yes         | Count matrix, spatial coordinates                                      | Expression patterns                                                    | R                    | https://github.com/MarcElosua/SPOTlight         |
Analysis of newly synthesised RNA has previously been limited to the level of cell populations, as weak detection tools required a large amount of total RNA. Dynamically produced RNA is an important measure of rapid gene expression in response to stimuli, a response which is not universal across cell populations. For this reason, a single-cell resolution technique is required. New transcriptome alkylation-dependent single-cell RNA sequencing (NASC-seq)\textsuperscript{72} is an emerging disruptive technology which monitors newly synthesised and pre-existing RNAs simultaneously (Fig. 1I), by sequencing the T–C chemical conversions in response to 4-thiothymidine (4T) labelling. This can be used to computationally separate, old and new RNA transcripts by sequence analysis, providing spatial and temporal monitoring at the single-cell level.

**Data visualisation and management**

There is high demand for public repositories of spatially resolved transcriptomic data, which allow comparison and integration of transcriptomic data, which allow comparison and integration of multiple sources and resolutions of analysed output\textsuperscript{73}. Progressing beyond curated databases, 3D-visualisation cell atlases modelling detailed spatial data will provide the most benefit in research and medical applications\textsuperscript{74}. Several applications exist which allow high fidelity imaging datasets to be integrated with quantitative data for completely interactive analysis, such as Virtual Fly Brain\textsuperscript{75}\textsuperscript{76}, and the atlas of embryonic human hearts which combines *Spatial Transcriptomics* and scRNA-seq\textsuperscript{90}. Results from *DistMap* were used to generate *Drosophila* Virtual Expression eXplorer (DVEX), an online, browsable database which generates virtual ISHs capable of predicting spatially resolved expression patterns in stage 6 embryos. Expression data is currently limited to binarized bins under defined thresholds, each expressing 6500–8500 genes, and cells are mapped independently with no incorporation of previously mapped cells into score refinement. Other 3D-visualisation tools such as MorphoNet\textsuperscript{76} will assist in integrating quantitative information from different platforms or sources on the same 3D object.

**Data management standards.** To promote reproducibility, standards for data storage, analysis and exchange are urgently needed, as spatial transcriptomics currently lack common guidelines for data management, compared to more mature sequencing technologies (e.g. RNA-seq, scRNA-seq)\textsuperscript{77}. Fortunately, the majority of analysis tools in spatial transcriptomics are publicly available on GitHub (e.g. <https://github.com/10XGenomics>, <https://github.com/SpatialTranscriptomicsResearch>), paving the way for building community standards, potentially enabling integration of different spatial transcriptomics technologies to address a common biological question. Further, it has become standard practice in the research community to have preprints publicly available (e.g. BioRxiv.org, MedRxiv.org), allowing rapid dissemination of methods and results, and lessons can be learned from the systems biology standards\textsuperscript{78,79}. This presents a great opportunity and a critical goal for the spatial transcriptomics community now that various technologies are in the developmental phase.

**Discussion and future directions**

Standard techniques such as in situ hybridisation are widely used to interrogate spatial gene expression. Based on visual examination, these techniques are extremely powerful in detecting changes in gene expression patterns. While providing useful data spatially, the quantitative power of fluorescence imaging remains low\textsuperscript{80}. Tools have continued to develop, with various improvements to the efficiency, versatility and power of these non-invasive imaging approaches\textsuperscript{81–83}. Recent improvements in single-molecule detection rendered this technology amenable for quantitative measurements and larger throughput, forecasting future scalability to assess whole transcriptomes.

RNA sequencing and its derivative technologies have answered the need for quantitative data and are suitably high-throughput for assaying whole genomes. These methods, however, lack the acuity necessary to deliver spatial results. Combining ISH and scRNA-seq using computational integration, achieves the level of accuracy required for resolving accurate spatio-temporal gene expression pattern. Identification of spatial markers and archetypal expression patterns is the basis for these current integrative approaches. Cell types can be visualised in 3D with computational techniques, however, this is based on measurement of spatial information from 2D slides and careful sampling is required to ensure faithful recapitulation of spatial patterning. Challenges remain in scaling these methods to accommodate large tissue volumes, however, as sequencing costs drop, it is forecast that it will be feasible for these technologies to scale to whole organisms. High-resolution, imaging-based approaches require high magnification and fine sectioning necessitating long imaging times, which will protract further with scaling. Methods which employ targeted probes carry the inherent limitation of finite available fluorophores, and will be challenging to scale. Underpinning these challenges is that, as more cells and tissues are analysed with higher resolution, more data points are required, demanding higher computational power, and scalable mathematical models for future 3D visualisation and data interpretation. Spatial technologies vary in their readiness for scaled analysis, and this will prove a key determinant of the lifespan and relevancy of the technology. Further limitations remain such as the difficulties in obtaining reproducible results due to single-cell analysis variability, and the applications in heterogeneous tissues such as eye retinas and cancer tumours remain to be determined. Methods to integrate spatial measurements with scRNA-seq data are beginning to emerge\textsuperscript{84}, employing “multimodal intersection analysis” to capitalise on the strength of information about cellular identity gained from scRNA-seq and the spatial data from microarrays to highlight spatially restricted gene networks and cell enrichment. Such approaches pave the way towards establishing virtual atlases which achieve a whole-organism transcriptome with cellular resolution. Several single-cell atlasing projects are endeavouring to develop these public resources which provide a valuable source of information for many model organisms including The Human Cell Atlas\textsuperscript{85} and single-cell zebrafish transcriptome atlas\textsuperscript{86}. Ultimately, spatially resolved transcriptomics will be paramount in adding spatial coordinates to these single-cell atlasing projects\textsuperscript{87}.

Accessibility to spatially resolved methods will determine popularity and speed of incorporation into research. Commercialisation of novel technologies supports consistent and reliable results, and promotes competition-driven enhancements to sensitivity and efficiency. CARTANA (Sweden) offers kits and services for padlock-based in situ sequencing technologies, while 10X Genomics acquired Spatial Transcriptomics and their Visium Spatial Gene Expression system reduced barcode spacing, leading to an improved resolution and 55 μm microarray\textsuperscript{88}. Ongoing modifications and novel data integration enhance utility, providing new versions of existing technologies such as Slide-seqV2 which is currently in development\textsuperscript{89}.

Several types of RNA molecules are active within the cell, engaging in diverse roles from mRNA encoding proteins, to non-coding and small RNAs regulating gene expression at the transcriptional, post-transcriptional and epigenomic levels\textsuperscript{90}. To generate a transcriptomic map capable of capturing the complexity of gene expression, it must be possible to detect a range of RNA molecules, unrestricted by length, and have the ability to distinguish between similar isoforms. Technologies which lack sequencing capabilities, also lack the ability to discriminate
between many alternate and splice forms, and cannot faithfully predict the effect of RNA structure. Approaches like smFISH which require multi-probe binding, restrict the discovery of short sequences of RNA such as microRNA and tRNA, and limit the profiling of different isoforms. Technologies which fail to capture subtle variations in sequence length and identity, or which lack the resolution to distinguish unique isoforms of RNA, comprise the complexity of these systems, and novel insights regarding cell type and gene function may be overlooked.

Rather than supplant existing techniques, many new approaches are designed to integrate with current technologies, expanding the scope of retrieved data. Many fields of science provide valuable insights into spatial genomics, however, without a wide-ranging, collaborative effort, these assets will remain underutilised. Mass spectrometry imaging is one such avenue yet to be fully explored for its integrative potential, despite demonstrating clear benefits in applications such as proteogenomics where structural imaging is followed by RNA sequencing. Computationally aligning spatially registered images with quantitative expression data will allow a unique appreciation of the complexity of tissue structure and composition. High-throughput spatial transcriptomics are becoming an essential component of both bench and bedside medicine, with an increasing demand for efficient analysis pipelines in clinical histology and pathology. The ultimate goal is to generate spatial-omics data, where a single sample may be non-destructively analysed to reveal temporal, spatial, and quantitative data simultaneously.

Integration of temporal expression datasets will enhance understanding of spatial transcriptomics. Incorrect temporal expression patterns can lead to developmental disorders or diseases like cancer. Understanding the gene expression timescale is crucial to identifying key triggers at different stages of development. For instance, single-cell studies performed at different stages of pluripotent cells differentiating into cardiomyocytes, revealed novel developmental pathways underlying lineage specification during cardiogenesis. Recent reports capturing expression data in vivo by tracking transcription factor binding have illustrated the possibilities of these techniques, including developing predictive GRNs. Integrating 3D regulatory information, such as epigenomics studies, will complete our appreciation of the GRN controlling spatio-temporal gene expression patterns. Advances in epigenomic studies to allow single-cell analysis, such single-cell ATAC-seq have provided new insights into regulatory heterogeneity in complex tissues. Ultimately, incorporating chromatin architecture and regulation in a 3D digital embryo will provide a comprehensive framework to uncover the GRNs executing spatio-temporal control of gene expression. This promises to accurately predict the downstream effects of genetic perturbations, such as those caused by mutations resulting in developmental disorders or disease, in an immediate, automated manner.

Received: 23 March 2020; Accepted: 10 September 2020; Published online: 23 October 2020

References
1. Holland, P. W. & Garcia-Fernandez, J. Hox genes and chordate evolution. Dev. Biol. 173, 382–395 (1996).
2. Mallo, M. & Alonso, C. R. The regulation of Hox gene expression during animal development. Development 140, 3951–3963 (2013).
3. Nelson, C. E. et al. Analysis of Hox gene expression in the chick limb bud. Development 122, 1449 (1996).
4. Zhang, M. et al. Ectopic Hoxa-1 induces rhombomere transformation in mouse hindbrain. Development 120, 2431–2442 (1994).
5. Reiter, J. F. et al. Gata5 is required for the development of the heart and endoderm in zebrafish. Genes Dev. 13, 2983–2995 (1999).
6. Driever, W. & Nüsslein-Volhard, C. A gradient of bicoid protein in Drosophila embryos. Cell 54, 83–93 (1988).
7. Saunders, J. W. Jr. Developmental control of three-dimensional polarity in the avian limb. Ann. N. Y. Acad. Sci. 193, 29–42 (1972).
8. van den Brink, S. C. et al. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature 580, 405–409 (2020).
9. Takasato, M. et al. Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. Nature 526, 564–568 (2015).
10. Chalá, G., Tyagi, S. & Ramalison, M. Navigating the non-coding genome in heart development and Congenital Heart Disease. Differentiation 107, 11–23 (2019).
11. DiPianto, D. J. et al. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. Thorax 70, 48–56 (2015).
12. Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia 27, 1000–1008 (2013).
13. Havelda, Z. & Maule, A. J. Complex spatial responses to cucumber mosaic virus infection in susceptible Cucurbita pepo cotyledons. Plant Cell. 12, 1975 (2000).
14. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176 (2007).
15. Bakken, T. E. et al. A comprehensive transcriptional map of primate brain development. Nature 535, 367–375 (2016).
16. Ding, S. L. et al. Comprehensive cellular-resolution atlas of the adult human brain. J. Comp. Neurol. 524, 3127–481 (2016).
17. Wang, Q. et al. The Allen mouse brain common coordinate framework: a 3D reference atlas. Cell 181, 936–953 (2020). The Allen Mouse Brain Common Coordinate Framework is an open access 3D reference atlas for the mouse brain at 10 μm voxel resolution providing a template for analysis, visualisation and data integration.
18. Harris, T. W. et al. WormBase: a comprehensive resource for nematode research. Nucleic Acids Res. 38, D463–D467 (2010).
19. Bowes, J. R. et al. Xenbase: gene expression and improved integration. Nucleic Acids Res. 38, D607–D612 (2010).
20. Lécuyer, E. et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell 131, 174–187 (2007).
21. Verbeek, F. J., Lawson, K. A. & Bard, J. B. Developmental bioinformatics: linking genetic data to virtual embryos. Int. J. Dev. Biol. 43, 761–771 (1999).
22. Jensen, E. Technical review: in situ hybridization. Anat. Rec. 297, 1349–1353 (2014).
23. Gawanlita, V. et al. Gene expression screening in Xenopus identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. Mech. Dev. 77, 141–151 (1998).
24. Kudoh, T. et al. A gene expression atlas in zebrafish embryo development. Genome Res. 11, 1979–1987 (2001).
25. Neidhardt, L. et al. Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. Mech. Dev. 98, 77–94 (2000).
26. Kukurba, K. R. & Montgomery, S. B. RNA sequencing and analysis. Cold Spring Harb. Protoc. 2015, 951–969 (2015).
27. Lignell, A., Kerosuo, L., Streichan, S. J., Cai, L. & Bronner, M. E. Identification of a neural crest stem cell niche by Spatial Genomic Analysis. Nat. Commun. 8, 1830 (2017).
28. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development. 145, https://doi.org/10.1242/dev.165753 (2018).
29. Wang, F. et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J. Mol. Diagn. 14, 22–29 (2012). RNAscope allows single molecule visualisation without the dissociation of tissue morphology and brings the advantages of ISH imaging to RNA biomarkers.
30. Rouhanifar, S. H. et al. ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. Nat. Biotechnol. https://doi.org/10.1038/nbt.4286 (2018).
31. Burke, K. S., Antilla, K. A. & Tirrell, D. A. A fluorescence in situ hybridization method to quantitify mRNA translation by visualizing ribosome-mRNA interactions in single cells. ACS Cent. Sci. 3, 425–433 (2017).
32. Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. Nat. Methods 15, 932–935 (2018).
33. Lyubimova, A. et al. Single-molecule mRNA detection and counting in mammalian tissue. Nat. Protoc. 8, 1743–1758 (2013).
34. Chen, H. K., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090 (2015).
35. Xia, C., Babcock, H. P., Moffitt, J. R. & Zhuang, X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. Sci. Rep. 9, 7721 (2019).

36. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc. Natl Acad. Sci. USA 116, 19490 (2019).

37. Vázquez, T. M. & Blank, C. C. U. A user’s perspective on GeoMxTM digital spatial profiling. Immuno-Oncol. Technol. 1, 11–18 (2019).

38. Ziai, J. et al. Abstract 2089: Highly multiplexed analysis of immune cell subsets using NanoString Digital Spatial Profiling. Nat. Methods 16, 987–990 (2019).

39. Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data. Nat. Commun. 10, 2084 (2020).

40. Quaife-Ryan, G. A. et al. Multicellular transcriptional analysis of mammalian embryos by MERFISH reveals subcellular RNA compartmentalization. bioRxiv https://doi.org/10.1101/2020.03.12.989806 (2020).

41. Rodríguez, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science 363, 1463–1467 (2019).

42. Wu, J. et al. Ribogenomics: the science and knowledge of RNA. Genomics Proteomes Bioinforma. 12, 709–722 (2019).

43. Rich, S. et al. Expression microarrays define cell type-specific spatial transcriptional proﬁling in the developing human brain. Cell 179, 1647–60.e19 (2019).

44. Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Sci Transl. Med. 10, 378–82 (2016).

45. Vázquez, T. M. & Blank, C. C. U. A user’s perspective on GeoMxTM digital spatial profiling. Immuno-Oncol. Technol. 1, 11–18 (2019).

46. Weinstein, J. A., Regev, A. & Zhang, F. DNA microscopy: optics-free spatio-genetic imaging by a stand-alone chemical reaction. Cell 178, 229–41.e16 (2019). BioRxiv https://doi.org/10.1101/2020.03.12.989806 (2020).

47. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Cell 177, 1873–87.e17 (2019).

48. Wolpert, T. et al. Comprehensive integration of single-cell data. Cell 177, 1888–902.e21 (2019).

49. Nizamic, M., Kariakin, N., Friedmann, M. T. & Rajewsky, N. Gene expression cartography. Nature 576, 132–137 (2019).

50. Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F. & de Wit, E. Systemic loss of neurofilamentous neurofilaments during zebrafish development. Proc. Natl Acad. Sci. USA 116, 755–760 (2019). bioRxiv https://doi.org/10.1101/2020.03.12.989806 (2020).

51. Karaiskos, N. et al. The Virtual Fly Brain browser and query interface. Nat. Methods https://doi.org/10.1101/2020.03.12.989806 (2020).

52. Combs, P. A. & Eisen, M. B. Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression. PLoS ONE 8, e71820 (2013).

53. Bryant, R. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptomes. Nature 576, 132–137 (2019). bioRxiv https://doi.org/10.1101/2020.03.12.989806 (2020).

54. Vasilevskas, G. et al. NASQ-seq monitors RNA synthesis in single cells. Nat. Commun. 10, 3138 (2019). NASQ-seq monitors newly synthesized and pre-existing RNAs simultaneously, providing single-cell, temporal analysis of the dynamic regulation of gene expression and RNA turnover.
Acknowledgements
We thank the members of the Ramialison Laboratory for their support. This project is funded by the Australian Research Council Discovery Project DP190102771 to M.R. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government.

Author contributions
L.N.W., L.G.M. and M.R. designed the study. L.N.W. wrote the manuscript with contributions from M.R., H.T.N. and L.G.M.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-01341-1.

Correspondence and requests for materials should be addressed to M.R.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020