

Application of Single-Cell Approaches to Study Myeloproliferative Neoplasm Biology



Daniel Royston, MBChB, BMSc, DPhil, FRCPath^a,
Adam J. Mead, MA, BMBCh, FTCP, FRCPath, PhD^{b,*},
Bethan Psaila, MA, MBBS, MRCP, FRCPath, PhD^b

KEYWORDS

• Myeloproliferative neoplasm • Single-cell genomics • Digital pathology

KEY POINTS

- Myeloproliferative neoplasms (MPNs) are an excellent tractable disease model for the application of single-cell approaches to study human disease.
- Single-cell genetic analysis of MPNs has provided important insights into disease latency and the importance of the order of mutation acquisition during disease pathogenesis, of broader relevance for cancer biology.
- Single-cell transcriptomics has revealed aberrant megakaryocyte differentiation trajectories in persons with myeloproliferative neoplasms.
- Digital pathology analysis combined with deep learning allows objective analysis of megakaryocyte heterogeneity in persons with MPNs.

INTRODUCTION

Philadelphia-negative myeloproliferative neoplasms (MPNs) are an excellent tractable disease model of a number of aspects of human cancer biology, including genetic evolution, tissue-associated fibrosis, and cancer stem cells.^{1,2} MPN is associated with long disease duration, well-characterized normal hematopoietic hierarchy, ability to purify cell populations by flow cytometry, and ease of accessibility of tissue derived from the malignant clone, facilitating the study of how this disease perturbs normal blood cell development through time. Accordingly, new technologies developed to study cancer biology have often been pioneered for the study of MPN. In this review,

^a Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine and NIHR Biomedical Research Centre, University of Oxford, Headley Way, Oxford OX39DS, UK; ^b Medical Research Council (MRC) Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, NIHR Biomedical Research Centre, University of Oxford, Headley Way, Oxford OX3 9DS, UK

* Corresponding author.

E-mail address: adam.mead@imm.ox.ac.uk

we discuss recent insights into MPN biology gained from the application of a number of new single-cell technologies to study human disease: single-cell genomics, single-cell transcriptomics and digital pathology (Fig. 1).

SINGLE-CELL GENOMICS

In the era of personalized medicine, genetic intratumoral heterogeneity (ITH) is increasingly recognized as a critical factor defining the behavior of a particular tumor in terms of clinical presentation, response to treatment, and risk of disease progression.³ Although bulk next generation sequencing (NGS) techniques have undoubtedly provided extensive insights into genetic diversity of clones within any given tumor,^{4,5} ultimately ITH can only be fully resolved using single-cell technologies.⁶ For example, inferring clonal structures from bulk variant allele fractions is inherently confounded by the presence of loss of heterozygosity or convergent evolution in which the same genetic events might occur multiple times in the same tumor.^{7,8} This makes it difficult to elucidate which mutations are present in the same clone, to accurately measure clonal diversity during therapy, to track disease evolution, or determine the order of mutations.

MPN has proven to be an important disease model providing an illustration of how single-cell genomics techniques can be applied to provide new insights into disease biology and how the technology might be moved toward application for precision medicine. Indeed, to a degree, single-cell genomics approaches are already in routine clinical use in myeloid diseases through application of cytogenetic analysis, including fluorescent in situ hybridization. The most common driver mutation in MPN occurs in exon 14 of the Janus kinase 2 (*JAK2*) gene, *JAK2V617F*, causing constitutive JAK-signal transducer and activator of transcription (JAK-STAT) signaling and driving the

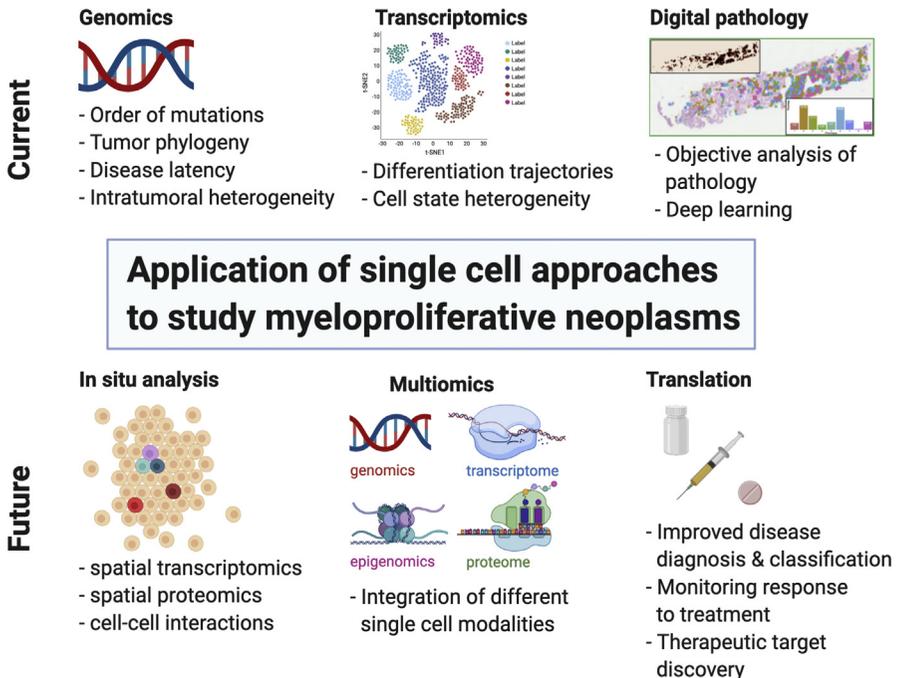


Fig. 1. Overview of single-cell approaches used to study MPN: current methods and future applications. Created with BioRender.com.

aberrant proliferation that is characteristic of MPN. Subsequently other mutations causing aberrant JAK-STAT signaling in MPN have been reported to occur in patients with MPN who are negative for the *JAK2V617F* mutation,^{9–11} confirming that constitutive JAK-STAT signaling is the key pathway driving MPN phenotype.¹² The mutational landscape of MPN has been extensively studied, revealing that MPNs are genetically relatively simple compared with other tumors.^{12,13} Together with other disease characteristics described in the introduction, this makes MPN an ideal tractable model for the application of single-cell genomics to understand the genetic evolution and clonal selection of mutant cells during disease development and progression. Consequently, single-cell assessment of genetic clonal evolution in MPN already has a rich history for the study of MPN over many years. Earlier studies benefited from the application of well-established clonogenic hematopoietic assays,² overcoming limitations of direct genetic analysis of single cells by providing an expanded cell population derived from a single cell for genetic analysis. For example, through the analysis of X-chromosome linked polymorphisms, including the study of single-cell-derived hematopoietic colonies, it has long been appreciated that MPNs are clonal diseases^{14,15} that develop on acquisition of a disease-initiating mutation in a single multipotent hematopoietic stem cell (HSC).¹ This mutant HSC clone undergoes a poorly understood process of clonal expansion over time, with subsequent proliferation of mature cells of the myeloid lineages.¹ At the time that the *JAK2V617F* mutation was first described,¹⁶ it was demonstrated that *JAK2V617F* selectively promoted the growth of single-cell-derived erythropoietin-independent colonies in patients with polycythemia vera (PV).¹⁷ This approach was also used to characterize clonal diversity in patients with MPN with concurrent *JAK2V617F* mutation and a cytogenetic abnormality, demonstrating that these genetic events were present in the same clone in some patients but in separate clones in others.¹⁸ Furthermore, combined analysis of *JAK2V617F* zygosity and loss of heterozygosity breakpoints using microsatellite markers in 6495 colonies revealed that *JAK2V617F* homozygous clones are recurrently acquired in patients with PV and patients with essential thrombocythemia (ET).¹⁹ However, PV was typically associated with presence of a dominant homozygous subclone, unlike in ET. The study of single-cell-derived colonies in MPN has more recently provided novel insights into the importance of the order of acquisition of driver mutations. Analysis of colonies derived from single cells from patients with JAK2 and TET2 co-mutated MPN revealed differences in disease phenotype and response to targeted therapy that were dependent on the order of acquisition of these mutations.²⁰ The acquisition of *JAK2V617F* first, followed by *TET2* mutation, was more likely to result in a PV phenotype, typically in younger patients, but if *JAK2V617F* was acquired on a *TET2*-mutated background (“*TET2*-first”), an ET phenotype was more frequent. Similar observations also applied to DNMT3A and JAK2 mutations.²¹ Together these studies nicely illustrate how the study of MPN can provide insights into pathways of genetic evolution of broader relevance for cancer biology.

Although analysis of hematopoietic colonies is a powerful approach, it is also associated with certain limitations and potential biases, as not all stem/progenitors are capable of growing colonies in vitro. Direct genetic analysis of single cells provides a potential solution, but is more challenging than analysis of material derived from colonies due to the very small amount of starting material and extensive amplification required.²² Single-cell mutation analysis can be carried out by targeted NGS of known mutations, single-cell exome, or whole-genome sequencing.²³ A number of methods have been developed, each with advantages and also limitations in terms of the spectrum of mutations that can be analyzed and the sensitivity and specificity of mutation detection.^{22,23} Whole-genome and whole-exome techniques allow the

characterization of new genetic events, whereas targeted techniques rely on detection of known mutations within a given tumor. Targeted mutational analysis is a higher throughput and more cost-effective approach than whole-exome or genome sequencing of single cells with improved sensitivity and specificity of mutation detection and high-throughput commercially available droplet-based platforms.²⁴ Whole-genome single-cell techniques are typically associated with higher rates of allelic dropout, although new approaches to study low-input genomic DNA look set to change this.^{6,25} To help address such technical problems, a number of bioinformatics tools to resolve the tumor phylogenetic trees have been developed.^{26–28}

One of the first examples of single-cell exome sequencing in human cancer was a case of JAK2-mutation negative MPN, where 58 cells were sequenced, revealing monoclonal evolution of the disease in this patient and potential new candidate mutations driving clonal evolution.²⁹ The integration of single-cell genotyping with fluorescence-activated cell-sorting purification of specific cell populations allows mutations to be mapped to distinct phenotypically defined cell types, confirming that all mutations in MPN can be tracked back to the phenotypic HSC population.³⁰ Single-cell targeted mutation analysis in serial samples from patients with myelofibrosis has revealed the very high level of clonal dominance in the CD34+ compartment seen in almost all patients with myelofibrosis.³¹ This study also elucidated pathways of clonal evolution during JAK2 inhibitor therapy, and demonstrated that complex clonal architecture correlates with risk of disease progression, particularly in association with the acquisition of RAS/RTK pathway mutations.³¹

New technologies now allow targeted mutational analysis of single cells to be done in a high-throughput manner using droplet-based approaches. For example, some of the first examples of direct single-cell sequencing using the Tapestry platform report the study of patients with MPN. In rare cases in which JAK2, CALR, and/or MPL mutations co-occur, this approach was used to determine that the mutations are present in independent clones.³² In a landmark study, the Tapestry platform has also recently been used to analyze 740,526 cells from 123 patients with myeloid malignancies, including cases of MPN that have progressed to secondary acute myeloid leukemia (AML).²⁴ As might be expected, the number of mutations present and clonal diversity was higher in AML than MPN and also higher in MPN than in clonal hematopoiesis. In 4 of 6 patients with transformation of MPN to secondary AML, a new dominant subclone emerged that in some cases was present as a minor subclone during chronic phase. In this study, the investigators also describe an exciting new methodology to combine protein expression at the single-cell level with genotyping to link phenotype and genotype.

Perhaps the most remarkable recent finding, revealed through single-cell genomics approaches, relates to the origins and disease latency of MPN. These studies used state-of-the-art single-colony whole-genome sequencing lineage tracing approaches that rely on identification of background somatic mutations as a “molecular clock” to determine the timing of clonal expansion and disease development following acquisition of the *JAK2V617F* mutation.^{33,34} With the caveat that both papers are available only as “preprints” and have not yet undergone peer review, it is striking that both studies reach a similar conclusion that the *JAK2V617F* mutation was reported to be acquired typically decades before disease development; remarkably, in many cases, the mutation was acquired in utero or in early childhood and yet only caused disease after many decades in adult life. In the study from the Cambridge group, 448,553 somatic mutations were identified and used to determine clonal dynamics in 843 hematopoietic colonies from 10 patients with MPN. This study estimated the median latency between *JAK2V617F* acquisition and disease onset to be 31 years, with

remarkable interpatient variation in fitness advantage of the MPN clone.³³ The study by Van Egeren and colleagues³⁴ analyzed a smaller number of colonies from 2 patients, also concluding that there was a disease latency of decades between *JAK2V617F* acquisition and disease onset. This long latency is particularly striking in view of the observation that many persons with normal hematopoietic parameters have evidence of a small *JAK2V617F* clone.³⁵ This suggests that many persons acquire a *JAK2V617F* mutation and live with this mutation for decades, and perhaps in many cases lifelong, without ever developing disease. The challenge is now to understand the heterogeneity of clonal fitness advantage exerted by the *JAK2V617F* mutation in HSCs. It is likely that this will involve an interplay among germline genetics influencing HSC biology,³⁶ heterogeneity of the HSC of origin, and extrinsic factors such as “inflammaging.”³⁷ To unravel this crucial aspect of MPN biology will no doubt require extensive use of single-cell methodologies that look set to be at the forefront of MPN research in coming years.

SINGLE-CELL TRANSCRIPTOMICS

Single-cell RNA-sequencing (scRNAseq) is the most widely applied assay in single-cell genomics, and is extensively used to provide a comprehensive and unbiased assessment of normal cellular and molecular tissue architecture and their perturbations in disease states. Over the past decade, experiments have massively expanded in their scale and implementation, due to technological advances resulting in high-throughput methods that are relatively easy to implement.^{38,39} In parallel, there now exists a wealth of user-friendly and open-source computational pipelines for data analysis.⁴⁰ Transcriptional profiling of cells individually has several advantages over “bulk” analyses, including detection of rare cell types; determination of whether differences between samples are due to differences in the frequencies of cell types present or alternatively changes in individual cell phenotype; and exploration of combinatorial patterns of gene expression and differentiation trajectories.

A typical analytical pipeline includes organizing cells according to their transcriptional profiles into discrete groups, or “clusters” that correlate with cell type or state.⁴⁰ Although cells are captured as transcriptional “snapshots,” their differentiation trajectories can be inferred computationally using trajectory analyses or ordering over “pseudotime,” to identify key transition states and bifurcation points.^{41–44} In addition, studying the ratio of spliced versus unspliced mRNA, or the “RNA velocity,” can be used to predict the future direction of travel of individual cells along a computed trajectory.^{45,46} scRNAseq can be readily combined with cell surface proteomics by incorporating barcoded antibodies,⁴⁷ and analytical techniques have been developed to infer cell-cell interactions using databases of receptor-ligand pairs.⁴⁸

scRNAseq techniques have been widely used to study MPN, as these diseases provide an exemplar model of a cancer involving complex interactions among malignant cells, diverse immune cell types (clonal and nonclonal), and mesenchymal stromal cells. In chronic myeloid leukemia, stem cells were studied from patients before and after tyrosine kinase inhibitor treatment and from the same patient before and after transformation from chronic phase to blast crisis.⁴⁹ Studying cells individually provided the necessary resolution to detect the rare, highly quiescent, BCR-ABL + stem cells in those responding to tyrosine kinase inhibition, and to demonstrate that these cells were transcriptionally distinct from normal stem cells, suggesting possible new targets for therapy. In this study, scRNAseq was combined with a novel method enabling BCR-ABL positive and negative cells to be reliably distinguished with high sensitivity, revealing that BCR-ABL negative stem cells also showed an aberrant

transcriptional signature with activation of inflammatory pathways, especially in those patients who failed to achieve an optimal response to treatment.⁴⁹

scRNAseq has also been applied to study how mutations alter hematopoiesis in Philadelphia-negative MPN, both in mouse models and in primary patient samples. In studies of primary cells isolated from patients with MPN, high-throughput scRNAseq has been applied to study how hematopoiesis is altered in patients with MPN.^{50,51} In a study of approximately 40,000 cells from patients with mutCALR-driven MPN, scRNAseq using a widely used droplet-based 3' scRNAseq platform (10x Genomics, Pleasanton, CA) with a new genotyping method involving targeted amplification of mutation transcripts was performed.⁵⁰ This method can easily be applied to profile 10 to 100s of 1000s of cells in parallel, and is sensitive for mutations that are highly expressed, such as mutCALR, for which approximately 90% cells were accurately genotyped, albeit less sensitive for low-expressed mutations (eg, 7.3% of JAK2V617F-mutant cells were genotyped) or those distant from the 3' end, for example, SF3B1 (~24% sensitive).⁵⁰ This study showed that CALR mutations in patients with ET affect the entire hematopoietic hierarchy, with mutant cells detected in all stem and progenitor subsets, although a higher proportion of mutant cells in the megakaryocyte progenitor (MKP) compartment. Trajectory analyses indicated that hematopoiesis was biased toward myeloid and myeloid-megakaryocytic differentiation, with MKP showing increased cell cycling, and mutant cells had upregulation of genes involved in the unfolded protein response to an NF- κ B signaling pathway.⁵⁰ A highly-sensitive method combining genomic DNA and complementary DNA genotyping in parallel with scRNAseq was also developed in MPN, enabling resolution of transcriptional signatures of genetic subclones in MPN and confirming that nonclonal stem/progenitors show aberrant, inflammatory gene expression signatures, highlighting the importance cell-extrinsic effects of MPN mutant clones.³⁰

A recent study of a mouse model of mutant calreticulin (CALR)-driven ET reported similar findings. In this model, mutCALR resulted in an expansion of both HSCs and megakaryocyte progenitors, and the investigators identified an aberrant intermediary population termed “proliferative megakaryocyte progenitors (pMKP)” that fell on a distinct differentiation pathway to MKP in normal hematopoiesis in their analyses.⁵² In addition, mutCALR HSC and MKP cell clusters showed significant dysregulation of genes involved in cholesterol biosynthesis as compared with wild-type cells,⁵² in addition to cell cycle and unfolded protein response genes as previously been described in mutCALR patients with MPN.⁵⁰

Profiling of more than 120,000 individual cells from a range of patients with primary and secondary myelofibrosis and both *JAK2V617F* and mutCALR-driven disease also demonstrated megakaryocyte-biased hematopoiesis, with an 11-fold increase in MKP detected in patients with myelofibrosis as compared with controls in all clinical and molecular subgroups.⁵¹ Notably, the MKP in patients with myelofibrosis fell into 2 distinct transcriptional subgroups: a small subset with a transcriptional profile similar to MKP detected in age-matched healthy donors, and a larger population with global upregulation of inflammatory/profibrotic genes. This study also identified that megakaryocyte-associated genes, including a cell surface marker, G6B, as being widely upregulated in myelofibrosis stem/progenitor cells, suggesting a strategy for immunotherapeutic targeting of cells derived from the myelofibrosis clone.⁵¹

Single-cell analyses have also shed light on the changes to the nonhematopoietic stromal cell compartment in MPN. Creating a “map” of certain stromal cell populations in myelofibrosis mouse models highlighted mesenchymal progenitor cells as showing the strongest upregulation in expression of extracellular matrix proteins in fibrosis. This study highlighted the S100A8/S100A9 alarmin complex as a potential therapy

target, demonstrating that its inhibition was able to ameliorate the disease phenotype in the mouse model.⁵³

These insights highlight the power of single-cell “omic” techniques to accurately dissect cellular and molecular perturbations in MPN. Combined with emerging techniques to capture cell states and transcriptomes in unperturbed tissues, so-called “in situ sequencing” (see [Fig. 1](#)) single-cell approaches will prove to be a powerful approach for target discovery and in the future look set to play a key role in clinical diagnosis through more accurate disease classification and risk stratification.

DIGITAL PATHOLOGY

The bone marrow represents a complex, dynamic, and highly regulated tissue in which diverse cell populations lie in close proximity to an orchestrated network of extracellular (stromal) matrix, blood vasculature, and bone.⁵⁴ This complexity is compounded by physiologic changes in response to aging, stress, and environmental factors that manifest as shifting patterns of tissue cellularity, heterogeneity, and lineage maturation.⁵⁵ Such complex spatiotemporal relationships are increasingly recognized as important for disease initiation, progression, therapeutic response, and relapse in patients with various myeloid malignancies, including MPN.⁵⁶ Although advances in high-resolution single-cell genomic technologies are well established in the search for new treatment strategies, advanced diagnostics, and disease monitoring in MPN, complementary approaches to decipher the important interactions among neoplastic hemopoietic cells, stromal constituents, and immune cell populations of the marrow in MPN are required.^{55,57} Recent developments in digital pathology, computer vision, and image analysis have the potential to address this imbalance and revolutionize the assessment of bone marrow tissues in MPN.

Key morphologic features relating to marrow cellularity, megakaryocyte pleomorphism/atypia, and fibrosis are firmly embedded in current MPN classification schemes.^{58,59} However, inconsistencies in the interpretation of key morphologic features may lead to inaccurate diagnosis and disease classification, with multiple studies suggesting significant intraobserver and interobserver variability among pathologists.^{60–63} Although this appears to be partly attributable to experience and training,⁶⁴ the subjective and qualitative nature of routine marrow biopsy reporting remains a fundamental limiting factor in any classification scheme incorporating morphology-based assessment of marrow tissue. The importance and value of more accurate and objective strategies for capturing the complexity of marrow tissue architecture in MPN extends beyond the potential for improving diagnosis and classification using current recommended criteria. As perturbations in the relationship between clonal and nonclonal hematopoietic cells and components of the marrow stem cell niche are gradually elucidated using sophisticated murine models of myeloid malignancies,^{55,56,65–68} therapeutic strategies targeting the mediators of tumor cell survival, proliferation, and chemoresistance are beginning to emerge.^{69–72} Translating these findings to human disease and validating novel therapies will require a concerted effort to move from the conventional, subjective, and laborious description of tissue morphologic features by pathologists to objective, quantitative, and automated descriptions of marrow constituents and their interactions.

Computational analysis of digitized images prepared from glass slide material has evolved over the past few decades and significantly accelerated in recent years with the development of sophisticated deep learning (DL) methods that emulate the structure and function of human neurons in the form of artificial neural networks.^{73,74} DL methods have found ready application in the field of pathology, with image

recognition convolutional neuronal networks (CNN) increasingly adopted for computer vision tasks in histopathology and cytology.^{75–77} In contrast to common solid tumors, image analysis has seen relatively limited application to disorders of the bone marrow, with most studies describing strategies for cell identification, quantification, and the resolution of specific leukemic differential diagnoses including B-ALL and common B-cell lymphoma/leukemia subtypes.^{78–83} These machine learning strategies have generally relied on morphology-based criteria to distinguish tumor cell subtypes rather than interrogate tumor cells with the intention of gaining novel insights into disease biology. However, recently machine learning has been used to correlate bone marrow aspirate morphologic features with somatic mutations in myelodysplastic syndrome, with specific morphologic profiles linked to unique clinical characteristics.⁸⁴

Despite the central role of bone marrow biopsy assessment in the diagnosis and classification of MPNs, and the importance of the marrow microenvironment in disease biology (as outlined previously), few studies have attempted to apply advanced machine learning approaches to these disorders. In response, we recently demonstrated the utility of an automated image analysis pipeline that uses machine learning techniques to extract important cytomorphological and topographic features of individual megakaryocytes from digitized images of bone marrow biopsies.⁸⁵ This enabled the differentiation of reactive samples from common MPN subtypes (ET, PV, and primary myelofibrosis) and assisted in disease classification. Clustering of megakaryocytes using the machine-learned features from extracted megakaryocytes identified cellular subtypes beyond the sensitivity of detection by specialist hematopathologists and were seen to correlate with the underlying MPN driver mutation status. When combined with topographic assessment incorporating patterns of megakaryocyte clustering and cell distribution, the extracted features could be combined to produce a multidimensional representation of an individual sample well beyond conventional microscopic assessment. Moreover, the rapid automated analysis of samples allowed index cases of MPN or reactive marrow to be contextualized against libraries of previously analyzed samples (Fig. 2). This could be used to monitor or track morphologic features over time, corresponding to either stable disease or

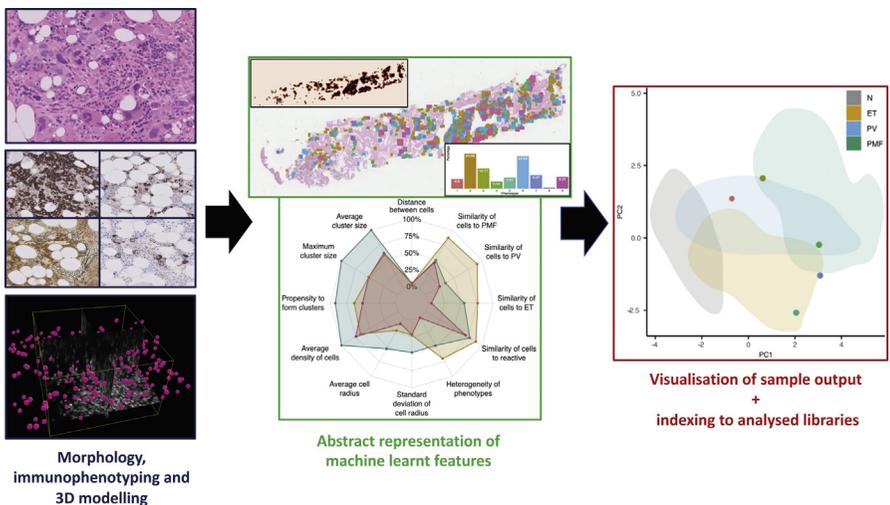


Fig. 2. Illustration of how digital pathology approaches can be used to study cellular heterogeneity in MPN.

progression. This work highlights the potential of image analysis, driven by advanced machine learning approaches, to improve tissue diagnosis in MPN and correlate tissue-based morphologic features with standard mutational and clinical data collected during the routine investigation of patients with MPN. Importantly, the automated extraction of objective quantitative data from routinely prepared hematoxylin-eosin-stained slides is ideally suited to future integration with the results of whole-tissue immunolabeling studies, advanced single-cell genomic analysis, and the outputs from high-resolution multiplexed tissue imaging performed in the research setting.

An important consideration in the development of improved descriptions of tissue morphology and the marrow microenvironment using digital pathology and machine learning is the development of intuitive yet sufficiently detailed data visualization. The methods used will depend on the application and requirements of the end user, but it seems likely that for clinicians and researchers unfamiliar with the normal bone marrow tissue architecture and its heterogeneity, visual representation of the cellular target(s) of interest in the context of normal and/or previously analyzed disease tissue will aid interpretation and understanding. Successfully designed outputs should capture temporal changes in the course of disease or following treatment, but also have the potential to highlight relevant diagnostic and prognostic features with attention drawn to potential therapeutic targets.

Notwithstanding the potential clinical application of automated analysis of discrete single-cell populations, such as megakaryocytes in biopsies of MPN, a deeper understanding of the complex cellular interactions within the bone marrow of patients with myeloid malignancies requires a more comprehensive description of the spatiotemporal relationships that exist between the cellular and stromal components of the marrow. This will require integration of discrete cellular and extracellular morphologic features with lineage specific markers of differentiation and maturation using multiplexed immunolabeling approaches. Although several analytical platforms already use such approaches in the research setting, they are typically restricted to relatively small tissue fields (roughly equivalent to conventional microscopic high-power fields) from limited sample numbers.^{86–88} Translating the insights of such studies to cohorts of routinely prepared clinical bone marrow samples will require the development of powerful and robust computational approaches that can be adapted to identify, quantify, and integrate diverse cellular and extracellular targets at scale. An additional challenge will be building advanced 3-dimensional models of the bone marrow environment and establishing methods for their validation using cohorts of patient samples analyzed in 2 dimensions.

Although automated analytical pipelines using convolutional neural networks to detect and segment targets of interest in digital images offer the potential for rapid sample analysis, their generation is typically dependent on access to large numbers of tissue samples accompanied by detailed clinical, laboratory, and genomic data. In general, machine learning applications in pathology use supervised approaches in which functions are learned by mapping annotated tissue features into some qualitative or quantitative output.^{77,89,90} This process is dependent on access to high-quality training data that are sufficiently labeled to allow the training phase to ultimately emulate the expert's input data. Strategies to reduce the burden of manual annotations by pathologists include transfer learning from preexisting CNNs and the development of human-in-the-loop annotation approaches that leverage human interactions to more rapidly train, test, and validate machine-learned functions.

Given the importance of accessing sufficient quantities of high-quality training and validation material, important practical considerations surround access to suitable

tissue. Few centers can rely solely on locally retained tissue archives to build and validate models of discrete MPN subtypes that are sufficiently enriched with relatively rare samples corresponding to important clinical or genomic events, such as disease transformation or response to novel therapeutics. Access to trial sample cohorts and sharing of tissue libraries between collaborating clinical centers will likely optimize use of available diagnostic material and accelerate the development and validation of machine learning models of MPN.

In summary, machine learning approaches to image analysis have already received broad acceptance in several branches of solid tissue pathology and are widely accepted as a transformational technology with significant clinical and research potential. Realizing this potential in MPN will depend on the identification and extraction of important cell-cell and cell-stroma interactions that complement and enhance our understanding of the dynamics underlying the clonal expansion of single-cell precursors that ultimately drive the disease phenotype in individual patients. This will require close collaboration among hematologists, pathologists, bioinformaticians, biomedical engineers, and software engineers and the integration of multimodality approaches spanning novel single-cell and whole-tissue sample technologies.

SUMMARY

Single-cell technologies have over many years provided remarkable insights into MPN biology, making conceptual advances of broader relevance across cancer biology. Future technological developments in digital pathology, in situ sequencing, and single-cell multiomics approaches (see [Fig. 1](#)) will undoubtedly be applied over the coming years to tackle crucial questions in the field. One key question that will definitively require single-cell approaches is why different HSC clones carrying MPN driver mutations such as *JAK2V617F* show such heterogeneity in fitness advantage. This is crucial to understand if we are to develop treatment approaches that reverse the fitness advantage to induce molecular responses and ultimately alter the natural history of MPN. Another key challenge in the field is to translate these single-cell technological developments through to direct patient benefit. Although at the present time it may seem farfetched for single-cell methodologies to be applied routinely in clinical diagnostics, this will surely become a reality over the coming years with obvious utility for improved approaches to diagnose and classify disease as well as monitor response to treatment and predict risk of disease progression.

FINANCIAL SUPPORT AND SPONSORSHIP

None.

ACKNOWLEDGMENTS

All authors contributed equally to this work. A.J. Mead is funded by a Cancer Research UK Senior Clinical Fellowship. This work was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Center (BRC). B.P. is supported by a CRUK Advanced Clinician Scientist Fellowship and Oxford BRC Senior Fellowship. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

DISCLOSURES

None.

REFERENCES

1. Mead AJ, Mullally A. Myeloproliferative neoplasm stem cells. *Blood* 2017;129(12):1607–16.
2. Nangalia J, Mitchell E, Green AR. Clonal approaches to understanding the impact of mutations on hematologic disease development. *Blood* 2019;133(13):1436–45.
3. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 2015;27(1):15–26.
4. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science* 2015;349(6255):1483–9.
5. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012;13(11):795–806.
6. Wills QF, Mead AJ. Application of single-cell genomics in cancer: promise and challenges. *Hum Mol Genet* 2015;24(R1):R74–84.
7. Baslan T, Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat Rev Cancer* 2017;17(9):557–69.
8. Brierley CK, Mead AJ. Single-cell sequencing in hematology. *Curr Opin Oncol* 2020;32(2):139–45.
9. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 2013;369(25):2379–90.
10. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 2013;369(25):2391–405.
11. Pardanani AD, Levine RL, Lasho T, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood* 2006;108(10):3472–6.
12. O’Sullivan J, Mead AJ. Heterogeneity in myeloproliferative neoplasms: causes and consequences. *Adv Biol Regul* 2019;71:55–68.
13. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;499(7457):214–8.
14. Adamson JW, Fialkow PJ, Murphy S, et al. Polycythemia vera: stem-cell and probable clonal origin of the disease. *N Engl J Med* 1976;295(17):913–6.
15. Gilliland DG, Blanchard KL, Levy J, et al. Clonality in myeloproliferative disorders: analysis by means of the polymerase chain reaction. *Proc Natl Acad Sci U S A* 1991;88(15):6848–52.
16. James C, Ugo V, Le Couedic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 2005;434(7037):1144–8.
17. Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005;365(9464):1054–61.
18. Beer PA, Jones AV, Bench AJ, et al. Clonal diversity in the myeloproliferative neoplasms: independent origins of genetically distinct clones. *Br J Haematol* 2009;144(6):904–8.
19. Godfrey AL, Chen E, Pagano F, et al. JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. *Blood* 2012;120(13):2704–7.
20. Ortman CA, Kent DG, Nangalia J, et al. Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med* 2015;372(7):601–12.
21. Nangalia J, Nice FL, Wedge DC, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica* 2015;100(11):e438–42.

22. Lim B, Lin Y, Navin N. Advancing cancer research and medicine with single-cell genomics. *Cancer Cell* 2020;37(4):456–70.
23. Povinelli BJ, Rodriguez-Meira A, Mead AJ. Single cell analysis of normal and leukemic hematopoiesis. *Mol Aspects Med* 2018;59:85–94.
24. Miles LA, Bowman RL, Merlinsky TR, et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* 2020;587(7834):477–82.
25. Ellis P, Moore L, Sanders MA, et al. Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nat Protoc* 2020. <https://doi.org/10.1038/s41596-020-00437-6>.
26. Roth A, McPherson A, Laks E, et al. Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat Methods* 2016;13(7):573–6.
27. Ross EM, Markowitz F. OncoNEM: inferring tumor evolution from single-cell sequencing data. *Genome Biol* 2016;17:69.
28. Sohrab Salehi AS, Roth Andrew, Aparicio Samuel, et al. ddClone: joint statistical inference of clonal populations from single cell and bulk tumour sequencing data. *Genome Biol* 2017. <https://doi.org/10.1186/s13059-017-1169-3>.
29. Hou Y, Song L, Zhu P, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 2012;148(5):873–85.
30. Rodriguez-Meira A, Buck G, Clark SA, et al. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol Cell* 2019;73(6):1292–1305 e1298.
31. Mylonas E, Yoshida K, Frick M, et al. Single-cell analysis based dissection of clonality in myelofibrosis. *Nat Commun* 2020;11(1):73.
32. Thompson ER, Nguyen T, Kankanige Y, et al. Clonal independence of JAK2 and CALR or MPL mutations in co-mutated myeloproliferative neoplasms demonstrated by single cell DNA sequencing. *Haematologica* 2020. <https://doi.org/10.3324/haematol.2020.260448>.
33. Williams N, Lee J, Moore L, et al. Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. *bioRxiv* 2020; 2020. 2011.2009.374710.
34. Van Egeren D, Escabi J, Nguyen M, et al. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in JAK2-mutant myeloproliferative neoplasms. *bioRxiv* 2020;2020. 2008.2024.265058.
35. Cordua S, Kjaer L, Skov V, et al. Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. *Blood* 2019;134(5):469–79.
36. Bao EL, Nandakumar SK, Liao X, et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* 2020;586(7831):769–75.
37. Karantanos T, Kaizer H, Chaturvedi S, et al. Inflammation exerts a nonrandom risk in the acquisition and progression of the MPN: insights from a Mendelian randomization study. *EClinicalMedicine* 2020;21:100324.
38. Stubbington MJT, Rozenblatt-Rosen O, Regev A, et al. Single-cell transcriptomics to explore the immune system in health and disease. *Science* 2017;358(6359):58–63.
39. Aldridge S, Teichmann SA. Single cell transcriptomics comes of age. *Nat Commun* 2020;11(1):4307.
40. Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet* 2019;20(5):257–72.
41. Wolf FA, Hamey FK, Plass M, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 2019;20(1):59.

42. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018. <https://doi.org/10.1038/nbt.4314>.
43. Coifman RR, Lafon S, Lee AB, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc Natl Acad Sci U S A* 2005;102(21):7426–31.
44. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32(4):381–6.
45. La Manno G, Soldatov R, Zeisel A, et al. RNA velocity of single cells. *Nature* 2018; 560(7719):494–8.
46. Bergen V, Lange M, Peidli S, et al. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 2020;38(12):1408–14.
47. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017;14(9):865–8.
48. Efremova M, Vento-Tormo M, Teichmann SA, et al. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;15(4):1484–506.
49. Giustacchini A, Thongjuea S, Barkas N, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* 2017;23(6):692–702.
50. Nam AS, Kim KT, Chaligne R, et al. Somatic mutations and cell identity linked by genotyping of transcriptomes. *Nature* 2019;571(7765):355–60.
51. Psaila B, Wang G, Rodriguez-Meira A, et al. Single-cell analyses reveal megakaryocyte-biased hematopoiesis in myelofibrosis and identify mutant clone-specific targets. *Mol Cell* 2020;78(3):477–492 e478.
52. Prins D, Park HJ, Watcham S, et al. The stem/progenitor landscape is reshaped in a mouse model of essential thrombocythemia and causes excess megakaryocyte production. *Sci Adv* 2020;6(48):eabd3139.
53. Leimkuhler NB, Gleitz HFE, Ronghui L, et al. Heterogeneous bone-marrow stromal progenitors drive myelofibrosis via a druggable alarmin axis. *Cell Stem Cell* 2020. <https://doi.org/10.1016/j.stem.2020.11.004>.
54. Mendez-Ferrer S, Michurina TV, Ferraro F, et al. Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* 2010;466(7308): 829–34.
55. Batsivari A, Haltalli MLR, Passaro D, et al. Dynamic responses of the haematopoietic stem cell niche to diverse stresses. *Nat Cell Biol* 2020;22(1):7–17.
56. Mendez-Ferrer S, Bonnet D, Steensma DP, et al. Bone marrow niches in haematological malignancies. *Nat Rev Cancer* 2020;20(5):285–98.
57. Dufva O, Polonen P, Bruck O, et al. Immunogenomic landscape of hematological malignancies. *Cancer Cell* 2020;38(3):424–8.
58. Guglielmelli P, Pacilli A, Rotunno G, et al. Presentation and outcome of patients with 2016 WHO diagnosis of prefibrotic and overt primary myelofibrosis. *Blood* 2017;129(24):3227–36.
59. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016; 127(20):2391–405.
60. Wilkins BS, Erber WN, Bareford D, et al. Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes. *Blood* 2008;111(1):60–70.
61. Alvarez-Larran A, Ancochea A, Garcia M, et al. WHO-histological criteria for myeloproliferative neoplasms: reproducibility, diagnostic accuracy and

- correlation with gene mutations and clinical outcomes. *Br J Haematol* 2014; 166(6):911–9.
62. Buhr T, Hebeda K, Kaloutsis V, et al. European Bone Marrow Working Group Trial on reproducibility of world health organization criteria to discriminate essential thrombocythemia from prefibrotic primary myelofibrosis. *Haematologica* 2012; 97(3):360–5.
 63. Harrison CN, McMullin MF, Green AR, et al. Equivalence of BCSH and WHO diagnostic criteria for ET. *Leukemia* 2017;31(7):1660.
 64. Barbui T, Thiele J, Gisslinger H, et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J* 2018;8(2):15.
 65. Arranz L, Sanchez-Aguilera A, Martin-Perez D, et al. Neuropathy of haematopoietic stem cell niche is essential for myeloproliferative neoplasms. *Nature* 2014;512(7512):78–81.
 66. Itkin T, Gur-Cohen S, Spencer JA, et al. Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* 2016;532(7599):323–8.
 67. Gleitz HFE, Dugourd AJF, Leimkuhler NB, et al. Increased CXCL4 expression in hematopoietic cells links inflammation and progression of bone marrow fibrosis in MPN. *Blood* 2020;136(18):2051–64.
 68. Christodoulou C, Spencer JA, Yeh SA, et al. Live-animal imaging of native haematopoietic stem and progenitor cells. *Nature* 2020;578(7794):278–83.
 69. Cho BS, Kim HJ, Konopleva M. Targeting the CXCL12/CXCR4 axis in acute myeloid leukemia: from bench to bedside. *Korean J Intern Med* 2017;32(2):248–57.
 70. Drexler B, Passweg JR, Tzankov A, et al. The sympathomimetic agonist mirabegron did not lower JAK2-V617F allele burden, but restored nestin-positive cells and reduced reticulin fibrosis in patients with myeloproliferative neoplasms: results of phase II study SAKK 33/14. *Haematologica* 2019;104(4):710–6.
 71. Ben-Batalla I, Schultze A, Wroblewski M, et al. Axl, a prognostic and therapeutic target in acute myeloid leukemia mediates paracrine crosstalk of leukemia cells with bone marrow stroma. *Blood* 2013;122(14):2443–52.
 72. Nimmagadda SC, Frey S, Muller P, et al. SDF1alpha-induced chemotaxis of JAK2-V617F-positive cells is dependent on Bruton tyrosine kinase and its downstream targets PI3K/AKT, PLCgamma1 and RhoA. *Haematologica* 2019;104(7):e288–92.
 73. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
 74. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform* 2016;7:29.
 75. Litjens G, Sanchez CI, Timofeeva N, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016;6:26286.
 76. Khosravi P, Kazemi E, Imielinski M, et al. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine* 2018;27:317–28.
 77. Rashidi HH, Tran NK, Betts EV, et al. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol* 2019;6: 2374289519873088.
 78. Kimura K, Tabe Y, Ai T, et al. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA. *Sci Rep* 2019;9(1):13385.

79. Singh A, Ohgami RS. Super-resolution digital pathology image processing of bone marrow aspirate and cytology smears and tissue sections. *J Pathol Inform* 2018;9:48.
80. Puigvi L, Merino A, Alferez S, et al. New quantitative features for the morphological differentiation of abnormal lymphoid cell images from peripheral blood. *J Clin Pathol* 2017;70(12):1038–48.
81. El Achi H, Khoury JD. Artificial intelligence and digital microscopy applications in diagnostic hematopathology. *Cancers (Basel)* 2020;12(4):797.
82. Alferez S, Merino A, Mujica LE, et al. Automatic classification of atypical lymphoid B cells using digital blood image processing. *Int J Lab Hematol* 2014;36(4):472–80.
83. Hagiya AS, Etman A, Siddiqi IN, et al. Digital image analysis agrees with visual estimates of adult bone marrow trephine biopsy cellularity. *Int J Lab Hematol* 2018;40(2):209–14.
84. Nagata Y, Zhao R, Awada H, et al. Machine learning demonstrates that somatic mutations imprint invariant morphologic features in myelodysplastic syndromes. *Blood* 2020;136(20):2249–62.
85. Sirinukunwattana K, Aberdeen A, Theissen H, et al. Artificial intelligence-based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients. *Blood Adv* 2020;4(14):3284–94.
86. Goltsev Y, Samusik N, Kennedy-Darling J, et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* 2018;174(4):968–981 e915.
87. Gerdes MJ, Sevinsky CJ, Sood A, et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci U S A* 2013;110(29):11982–7.
88. Di Palma S, Bodenmiller B. Unraveling cell populations in tumors by single-cell mass cytometry. *Curr Opin Biotechnol* 2015;31:122–9.
89. Chang HY, Jung CK, Woo JI, et al. Artificial intelligence in pathology. *J Pathol Transl Med* 2019;53(1):1–12.
90. Serag A, Ion-Margineanu A, Qureshi H, et al. Translational AI and deep learning in diagnostic pathology. *Front Med (Lausanne)* 2019;6:185.