



QuPath: The global impact of an open source digital pathology system

M.P. Humphries^a, P. Maxwell^a, M. Salto-Tellez^{a,b,*}



^a Precision Medicine Centre of Excellence, The Patrick G Johnston Centre for Cancer Research, Queen's University, Belfast, UK

^b Integrated Pathology Programme, Division of Molecular Pathology, The Institute of Cancer Research, London, UK

ARTICLE INFO

Article history:

Received 19 October 2020
Received in revised form 11 January 2021
Accepted 14 January 2021
Available online 21 January 2021

Keywords:

QuPath
Digital Pathology
Computational Pathology
Image Analysis
Artificial Intelligence

ABSTRACT

QuPath, originally created at the Centre for Cancer Research & Cell Biology at Queen's University Belfast as part of a research programme in digital pathology (DP) funded by Invest Northern Ireland and Cancer Research UK, is arguably the most widely used image analysis software program in the world. On the back of the explosion of DP and a need to comprehensively visualise and analyse whole slides images (WSI), QuPath was developed to address the many needs associated with tissue based image analysis; these were several fold and, predominantly, translational in nature: from the requirement to visualise images containing billions of pixels from files several GBs in size, to the demand for high-throughput reproducible analysis, which the paradigm of routine visual pathological assessment continues to struggle to deliver. Resultantly, large-scale biomarker quantification must increasingly be augmented with DP. Here we highlight the impact of the open source Quantitative Pathology & Bioimage Analysis DP system since its inception, by discussing the scope of scientific research in which QuPath has been cited, as the system of choice for researchers.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	852
1.1. The application of QuPath	853
1.2. The need for reliable analysis in a digitised framework	853
2. Methodology	854
2.1. The spread of QuPath	854
2.2. QuPath, a community	854
2.3. The impact of QuPath	855
2.4. The flexibility of QuPath	856
2.5. Abstract and outlook	856
Funding	857
CRediT authorship contribution statement	857
Declaration of Competing Interest	857
References	857

1. Introduction

The use of open source software is becoming a key component of modern scientific activity. Indeed, there is increased evidence

that some of the key discoveries in many areas of science would have not been possible without open source tools [1]. Of the thousands of scientific applications world-wide, the use of open practices and open resources in the field of digital pathology has revolutionizing tissue-based image analysis [2]. In areas such as cancer diagnostics and cancer research, there is an increasing interest in analyzing how these practices are dictating patient management and patient stratification [3]. We hereby analyze

* Corresponding author at: Precision Medicine Centre of Excellence, The Patrick G Johnston Centre for Cancer Research, Queen's University, Belfast, UK.

E-mail address: m.salto-tellez@qub.ac.uk (M. Salto-Tellez).

how QuPath, arguably the most widely used image analysis software in the world, has impacted the quantitative analysis of tissues and cells in research and diagnostics, as a way to illustrate how these tools are influencing the delivery of contemporary research.

QuPath, short for Quantitative Pathology, is an open source software with an active and engaged community able to support the development of tools for image analysis. The platform allows researchers with software development skills to add their own extensions to solve new challenges, but was designed for users without computer programming skills. QuPath is able to exchange data with existing tools such as ImageJ and MATLAB, and while ImageJ is perhaps the best known open source software for biomedical image analysis, it has historically struggled to deal with the size of whole slide images (WSI). On the back of the explosion of DP and a need to comprehensively visualise and analyse WSIs, QuPath was developed to address the many needs associated with tissue-based image analysis [4]. These needs were several fold and predominantly translational in nature: from the requirement to visualise images containing billions of pixels from files several GBs in size, to the demand for high-throughput reproducible analysis, which manual pathological assessment continues to struggle to deliver [5–7]. Resultantly, large-scale biomarker quantification must increasingly be augmented with DP.

QuPath was designed with the user in mind, with an easy-to-use interface that provides researchers and diagnosticians an ability to easily navigate complex tasks, such as automatically detecting many thousands of objects, classify and count these across large images, all while providing flexibility and high-throughput automated processing capabilities. Specific documentation on the introduction to analysis, installation, useful tutorials and much more is maintained and curated by the lead creator and developers of QuPath and can be found here: <https://qupath.readthedocs.io/>.

Due to its comprehensive nature as a tool for working with WSI, QuPath is a cross-platform software application designed for bioimage analysis that can be applied to numerous types of images beyond pathology. Although primarily created and used in cancer research for high throughput biomarker analysis in immunohistochemically stained formalin fixed paraffin embedded tissues, QuPath is able to meet the needs of many users: from laboratory researchers wishing to obtain raw quantitative data, to computational scientists working on the development and testing of algorithms. Primarily, QuPath software was designed for WSIs in digital pathology, to enable the analysis of immunohistochemistry (brightfield or fluorescent) and haematoxylin and eosin (H&E) images. QuPath's interactive interface is user-friendly and is able to count cells, and classify objects and pixels in large WSIs without the need for cropping or down-sampling images to lower file sizes for subsequent analysis.

Herein, we highlight the global impact of QuPath since its inception and discuss the scope of scientific research in which QuPath has been cited as the system of choice for researchers. Lastly, we aim to illustrate the need for reliable analysis in a digitised framework especially with regard to the development of artificial intelligence (AI) models.

1.1. The application of QuPath

Detailed information, including specific resources for support, documentation, compatible file types and instruction on the use of QuPath are available <https://qupath.github.io/>, <https://qupath.readthedocs.io/>, <https://www.youtube.com/c/QuPath> and are considered beyond the scope of this article. Broadly speaking, the software can support a wide variety of applications. QuPath is able to open and present WSIs from a wide range of file type's from many digital pathology scanners. It is perhaps important to highlight that QuPath does support the Digital Imaging and Communications in

Medicine (DICOM) standard format for WSIs. Users can view associated properties of an image, and using intuitive controls and gestures, navigate around images, panning and zooming as needed. Users can easily annotate tissues manually or with automated processes. Detection of cells is effortless with in-built steps and helpful default settings. Users can simply view measurements and export data. While performing analysis in this interactive step-wise method has its uses, the software contains the potential to create and apply similar analysis in a reproducible batch processing manner across a large image sets.

The flexible WSI viewer, which incorporates specific tracking slide navigation, is the cornerstone of many other relevant tools [8]. QuPath supports the quantification of hybridization signals with specific subcellular localization detail, automated tumour identification, tile-splitting of images to encourage faster analysis, estimation of stain intensity, capacity to exchange data with open source tools and the ability to read many image types, while carrying out real-time analysis with clinic-pathological parameters.

These automated processes are easily arranged into simple workflows created from the command history. As an example of a popular workflow within QuPath, the assessment of tissue microarrays (TMA) begin with the creation of an analysis project followed by automated de-arraying of the TMA and estimation of staining. A single-cell detection is then undertaken, followed by trainable cell classification. Following biomarker detection, data may be further analysed within the software or exported. QuPath also facilitates batch processing, and contains comprehensive survival analysis tools. Creation of workflows, allows the possibility to automate some or all of the analysis across multiple images. Viewing the command history allows for the creation of a script containing all the steps which have thus far been applied to an image. QuPath can be instructed to run a script over all project images to be analysed in a reproducible way. The flexibility of QuPath is continually evolving and scripting within QuPath can speed up analysis considerably. Advanced users recognise that scripting is able to unlock a huge range of features and possibilities.

1.2. The need for reliable analysis in a digitised framework.

The need for reliable analysis in a digitised framework is applicable to research and clinical applications, including diagnosis [9] and prognosis [10]. The increasing interest in digital pathology as a tool that supports discovery and delivers clinical utility may have arisen from the series of technical improvements which have transformed DP over the last 40 years into a reliable application [11]. These improvements and transformations include high-quality whole slide scanners, adequate image sizing and compression, speed and automation of the scanning process, and numerous options for image storage, and real-time retrieval/review. These technical developments have enabled DP to be the conduit to bring quantitation to a paradigm previously dominated by subjectivity. Quantitation such as: measuring nuclear morphology and DNA content; quantitative immunohistochemistry; analysis in multiple formats (e.g. tissue microarrays, appreciation of tumour heterogeneity and phenotypic variation); and capacity to identify and quantitate both chromogenic and fluorescent signals is swiftly achieved using DP. Resultantly, DP is able to overcome one of the main challenges we face in tissue analysis today, namely quantitative biomarker discovery. Indeed, DP can in parallel support the drug development and associated biomarker analysis process [11].

In the research environment, it is now clear that the need for more complex image analysis is growing exponentially and is transforming tissue-based discovery. These needs include an evaluation of broad topographic components of disease; for instance, the tumour stromal ratio in cancer samples or the characterization of tumour budding at the epithelial-mesenchymal transition [12].

A clear example of the need for image analysis and the application of deep learning to extract sub-visual features which can inform our understanding of geospatial variability of host immunity, is typified by the work of AbdulJabbar et al. [13]. In this study, the authors created a deep learning pipeline that enabled the spatial mapping of four distinct cell types within H&Es: malignant epithelial cells; lymphocytes; and non-inflammatory stromal cells, with an additional class of other non-identifiable cells or less abundant cells. This workflow collectively elucidated immune-evasive tumour microenvironmental mechanisms that may permit the emergence of aggressive clinical phenotypes [13]. The complexity and importance of the prognostication of spatial relationship is being further elucidated within the literature [14,15]. Recently, Tsakiroglou et al. concluded that the proximity (<30 µm) of CD8 t-cells to PD-L1+ cells, as well as PD-1+ cells to PD-L1+ cells, was prognostic for overall survival in patients with head and neck squamous cell carcinoma. The authors cite QuPath's custom scripting capability, active supportive community, well-maintained documentation, version management and open source nature as rational for utility with in their study.

Through the quantitation of tissue-hybridisation signals for specific biomarkers to confirm prognostic molecular pathways [10], and diagnostic solutions [16], we have shown how image analysis can greatly inform tissue-based discovery. Taking a robust digital pathology approach, we evaluated the suitability of alternative immune classifiers on prognostication in two independent colorectal cancer cohorts. We went on to assess the relationship with biology amenable to targeted therapy in a clinical trial cohort [10]. We have additionally shown the application of image analysis beyond discovery into a clinical application is not only feasible but much needed in the improvement of PD-L1 diagnostic accuracy [9,16]. More generally, the application of image analysis and specifically deep learning, may be a prerequisite in the predication of molecular and outcome data from simple H&E images.

Sirinukunwattana et al. were able to take complex tissue organisation features from unclassifiable or heterogeneous cases of colorectal cancer and accurately predict RNA expression [17]. While not yet clinically viable, increasing the level of actionable data extracted from routine H&E slides, inaccessible to human interpretation, may providing vital information regarding tumour heterogeneity, with application potential in settings where previously there may have been no means of applying expensive molecular testing.

In the histopathological diagnostic setting, the need for digitisation and DP implementation cannot be overemphasised. Key elements to support such diagnostic decisions include [18]:

1. The need for accurate biomarker analysis in leading reference hospitals. Clinical trials in tertiary healthcare are in urgent need of digitisation, and in many cases are a *conditio sine qua non* for trial delivery
2. The need for reproducibility, consistency and accuracy in phenotypic diagnostics
3. The relevance of DP as a cost-effective tool, with formal accreditation by quality agencies
4. The need of a digitised service to apply AI solutions
5. The requirement for solutions allowing remote pathology diagnosis in the context of pandemics

It is in this fertile ground that we believe the numerous advantages of QuPath have flourished to support basic, translational and clinical research.

2. Methodology

The original paper [4], cited anywhere between 400 and 679 times at the time of writing according to a Web of Science Core Col-

lection [19], and Google Scholar Cited Reference Search, spans many disciplines. Due to the prevalence of pre-print articles returned via Google Scholar, which are defined as preliminary reports that have not been peer-reviewed, we have chosen to focus on the conservative citation number from Web of Science, with the caveat that QuPath has been utilised in many more publications than cited. Our data does not include the frequent use of QuPath in abstracts for major scientific conferences such as USCAP and AACR meetings which we acknowledge are numerous.

2.1. The spread of QuPath

The areas of specialization are numerous and span a broad range of disciplines from oncology to computer science, although it is perhaps unsurprising that oncology, cell biology and pathology are the more common areas of research that regularly utilise QuPath, as shown Fig. 1.

Of these disciplines, there are >50 countries and regions represented that have published experience with QuPath in terms of the institutional affiliations associated with publications, (Fig. 2). Many of these articles associate with multiple countries for which the argument could be made that not only has DP facilitated and fostered national and international inter-laboratory collaborations, but owed to its open-source nature, that QuPath itself has connected institutions which otherwise would have little means of collaboration.

The utility of QuPath continues to grow unabated year on year, with publications growing at a rapid pace (Fig. 3). A cursory search of PubMed reveals that each year more authors and institutions are utilising QuPath for their research, although citation of the primary source is sometimes lacking, e.g. 2017–2020 [20–23].

In addition to published research, communication and knowledge gathered internally and throughout our CRUK Accelerator network (<https://www.qub.ac.uk/research-centres/PMC/cruk-centres-network-accelerator-award/>) and beyond, indicate that QuPath is routinely used for the scoring of stained biomarkers in clinical trial material. Indeed, we are aware of contract research organisations, charities and biopharma are beginning to have an interest in utilising QuPath in their data analysis pipelines. The impact of QuPath is seen not only across domains as indicated above, but across translation boundaries within specialisms. The application of QuPath is regularly seen at the basic research level [24,25], also within the translational research sphere [10,26,27], and recently in the clinical domain [16,28]. Our own experience with contract research has indicated that in addition to off-the-shelf image analysis solutions, requests for biomarker analysis increasingly specify the use of QuPath for quantification <https://www.qub.ac.uk/research-centres/PMC/Filestore/Fileupload,972238,en.pdf>. From a cost-analysis perspective, the ability to utilise an open-source DP platform with non-inferior performance to off-the-shelf DP solutions is appealing. In one such inter-platform and inter-operator analysis using the proliferation marker Ki67, a key comparative study concluded that QuPath was indistinguishable from others [29].

2.2. QuPath, a community

Owing to the open-source nature of QuPath, motivated and like-minded individuals can come together to contribute and share their experiences of QuPath. User forums provide a conduit to share workflow scripts, opinions and plans with other community members, general users and users of other complementary valuable open source projects.

The dialogue between users serves as a platform for development of novel algorithms, workflows and workarounds (<https://forum.image.sc/tags/qupath>), as well as the identification of software



Fig. 1. QuPath Research Areas. A treemap visualisation of the 15 top research areas assigned to the publications citing the seminal QuPath paper. Publications are not restricted to a single discipline.

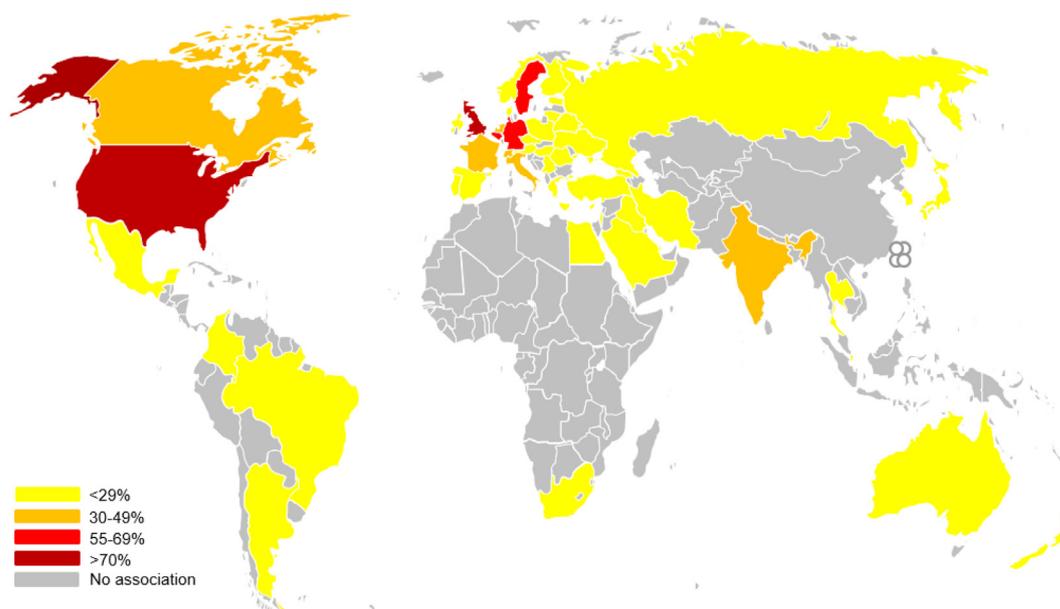


Fig. 2. The Global Reach of QuPath. A visualisation of the regions and countries with institutional recognition of the use of QuPath, by way of publication.

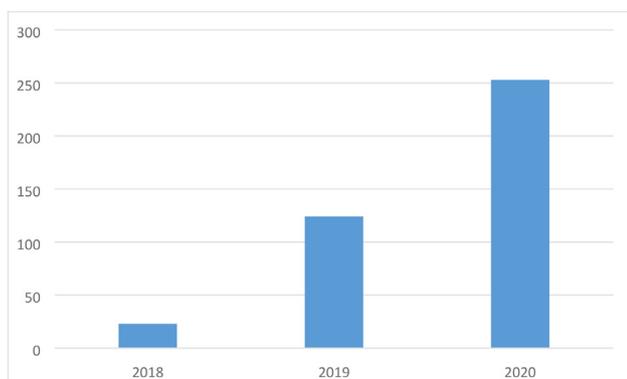


Fig. 3. The Citation of QuPath Per Year. Web of Science Core Collection Cited Reference Search. Data last accessed 31/12/2020 11:00.

“bugs” or glitches (<https://github.com/qupath/qupath/issues>). At a time in which intellectual property and software development appear to be an indispensable component in developing a reliable research tool, QuPath’s co-operative model provides a fresh alternative. A high-quality academic open-source development, sustained and improved by a community of users represents a pathway with many advantages such as: accessibility, affordability and access to constant development of the original published platform. This ecosystem is at the heart of the ongoing evolution of QuPath.

2.3. The impact of QuPath

QuPath’s adoption by research groups delivering highly impactful research is undeniable. The ability to reliably and reproducibly quantify biomarkers has enticed researchers to use QuPath from round the world. As demonstrated by its appearance in some of

the most impactful publications in journals such as Nature [30–32], Cell [33–36] and Science [37,38]. Roberti et al. used QuPath to quantify the cell density of DAB-positive cCasp3 cells; these data in part contributed to the demonstration that immunogenic ileal apoptosis contributed to the prognosis of chemotherapy-treated colon cancer [30]. The ability to use QuPath to robustly quantify Ki67 is well documented [29,39]. In one Nature study the authors used QuPath to quantify the frequency of Ki67 positive cells in the whole crypt region of small intestine samples; these data contributed to the revelation of a regulated dynamic neuro-immune circuit where a trade-off between innate immune protection and efficient nutrient absorption was found. Indeed, this mechanism may be effective for enhancing resistance to pathogens and in the treatment of metabolic diseases [31]. A less often published capability of QuPath is the ability to robustly quantify BaseScope and RNAScope, two applications which allow for the *in situ* visualisation of biological functional units such as cell junctions, and RNA. Again in Nature, QuPath was used in the field of multiple sclerosis to quantify the number of BaseScope-positive signals in annotated regions. These data contributed to evidence that oligodendroglial heterogeneity in multiple sclerosis may be important for understanding disease progression and developing therapeutic approaches [32].

In developing the prognostic use of QuPath as a tool for annotation training for deep learning, Liu et al. [40] showed in 843 samples that their network could stratify patients with nasopharyngeal carcinoma into a high risk group with shorter than 5 year progression free survival ($p < 0.0001$).

It is clear that the majority of publications have used QuPath in brightfield histopathological assessment of biomarkers in FFPE sections, be these in resections [39,41], biopsies [42,43], cytology specimens [9,44], TMA [10,45–48], or embedded cell culture models [49,50]. This is in addition to the multitude of immunofluorescence applications of QuPath [14,16,36,51,52]. Yet the intuitive annotation and stain quantification capabilities have enabled the application of QuPath in other, perhaps unexpected domains. For example, QuPath was used to perform semi-automated boundary detection of particles from X-ray fluorescence images from the mining domain, i.e. the processing of extracting precious mineral resources such as gold and silver [53]. Here the authors used pixel counting and boundary detection using specific colour values in QuPath to estimate the optical density of relatively high-arsenic regions in images which are critical to assessing potential short- and long-term exposure health risks to humans and surrounding environments.

2.4. The flexibility of QuPath

QuPath has undoubted utility in the quantification of ‘routine’ DAB biomarkers [9,10,54–57], including RNAScope [28,58–60], and the ability to handle complex quantification of multiple immunofluorescence biomarkers [16,23,51,61]. As we have seen in the examples above, QuPath provides a framework for the training, provision and application of advanced AI, beyond that of the inbuilt machine learning methodologies. This framework could consist of anything from the training of more advanced deep learning neural networks by way of pathologist annotation or patch extraction, to ultimately the visualisation of the deep learning models trained on data acquired outside QuPath [62–67].

Using open source software for research purposes is useful, but translating such research to clinical utility poses challenges. A means of software version control is essential whereby changes in software versions are tracked, maintaining an understanding and traceability of how iterative versions of the same software differ one from another. In this way, regulatory bodies when reviewing software developments for use in clinical workflows can

understand the mechanism by which a clinical decision is made, or trace any occurrences of clinical error, ensuring sufficient data for root cause analysis. Such measures required for the development of AI tools in clinical trials have been outlined [68]. The use of QuPath in the development of AI via deep learning techniques in a supervised or semi-supervised manner requires a level of version control to meet these challenges. The integration with a quality management system whereby documentation and version control is maintained, along with fully curated sample cohorts and robust metadata is matched with training and competency records of laboratory staff and those providing annotations. Independent and ongoing review of annotation data and feedback between data scientist and reviewer can identify any need for retraining of annotators (Fig. 4). In such a process version control of developing networks is possible. Tracking the version of QuPath used for annotations is recorded for the training and testing of neural networks, therefore resulting models can be version controlled, even when using open source software.

It is important to highlight here the need for widespread adoption of data version control (DVC) in the development and collaboration of AI models. DVC is a data versioning and experimental management tool, which builds upon established workflows. The sharing and collaboration often done through a standard Git-flow (i.e. commits, pull requests, etc.), can be combined with DVC to enable data scientists and machine learning teams to version control experiments, make projects reproducible and curate large datasets [69]. In addition to the accessibility of image training datasets and methods of classification, the quality of coding in AI and data driven machine learning applications is a pre-requirement for successful and sustainable software development. Property-based testing methodology ensuring code quality is necessary for the quality assurance, interpretation and tractability of graphical learning models [70]. These factors, as well as specifying programming languages utilised and clarifying whether these process are fully or semi-automated, need to be considered with a holistic appreciation for the complexity of a pathology department. While the specifics of any one element are vital decisions which must be taken for the introduction of AI and machine learning, an end-to-end overview is necessary to appreciate the interdependencies of all workflow components [71].

As the third revolution in pathology evolves [72], the need for robust, and reliable bio-imaging tools will become necessary, and one where users of QuPath can rise to meet the challenge (Fig. 4).

2.5. Abstract and outlook

Disruptive technologies are often embraced by many, while simultaneously resisted by those whose faith in the prevailing paradigm is based on years of tried and tested methodologies. Since its inception, QuPath has achieved a remarkable uptake in its use across the domain of pathology. However, a limitation of our review of this space fails to capture the use of QuPath beyond that collated within publication repositories, such as the impact of QuPath within pathology education or the specific number of users within the biopharma industry. More recently, the flexibility of the system has appealed to researchers from other disciplines. Within a discovery research setting, it is clear that image analysis, and the application of deep learning has clear utility and robustness. The challenges facing clinical pathology is whether a seamless integration of image analysis into a digitised service is not only reproducible, consistent and accurate, but that it can be user-friendly, cost-effective and fold seamlessly into the framework of accreditation laboratories, while adding substantive value in supporting clinical services. Furthermore, improvements in explainable and transparent AI away from the perception of a black-box is essential for such models to be not only usable but also useful to the expert.

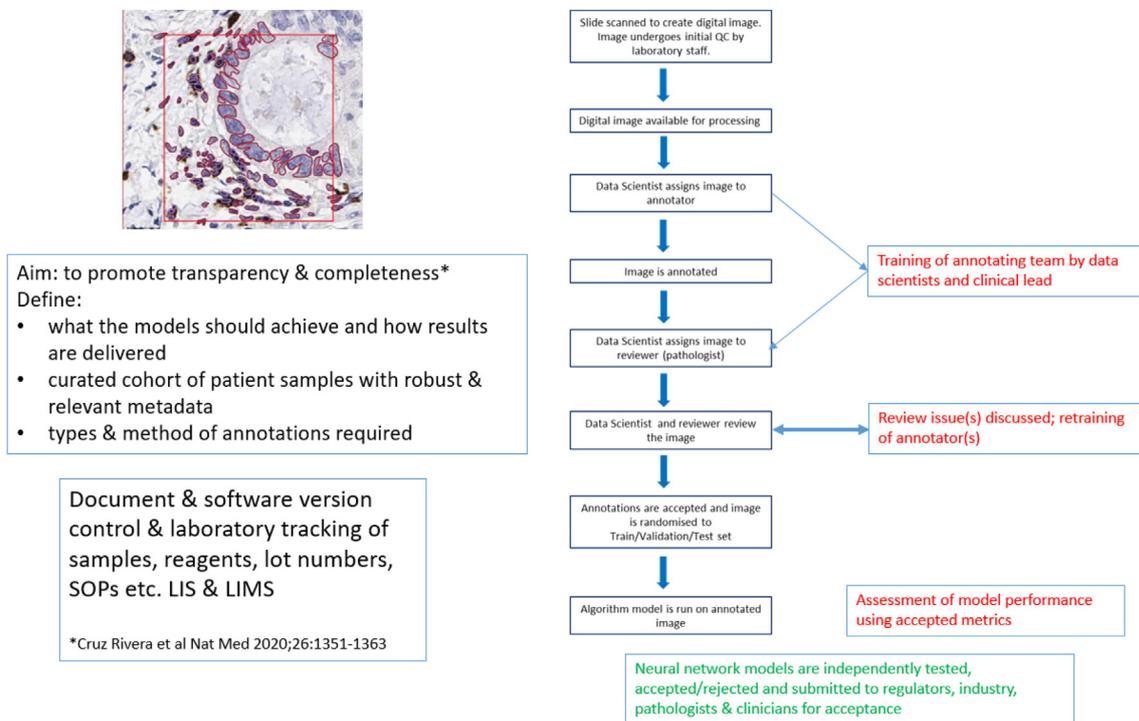


Fig. 4. Translating research to practice. Workflow outline of using open software such as QuPath as a training tool for deep learning neural networks within a quality management system which maintains document and software control. Alongside training and competency of laboratory staff, annotators are trained and output is independently reviewed before being used in the training of the network. Change control is thus recorded to aid in the understanding of the development of the neural network for acceptance by regulators and the clinical community.

The usability of AI models will build trust overtime allowing measurement of effectiveness, efficiency and satisfaction for users [73]. Whether QuPath may play a role in this future remains to be seen.

The concept of using open-source software either as stand-alone machine learning quantitative tools, or as a means to support the development of AI for clinical deployment will need to be used within a version controlled quality management system. An environment which can be validated and critically evaluated by regulators, pharma and co-industrial partners, and the clinical community. Commercial platforms are therefore seen as having the support from commercial-grade & quality assured frameworks and are perhaps better positioned to support image analysis for clinical utility. What is in no doubt, is that the development and application of image analysis and the development of AI will continue unabated. The translation of these tools from a research setting into a clinical setting through acceptance by regulators, industry and clinical groups is key to driving the next revolution in tissue based biomarker discovery.

The latest release of QuPath is accessible here: <https://qupath.github.io/>

Funding

Funding information is not applicable.

CRediT authorship contribution statement

M.P. Humphries: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. **P. Maxwell:** Conceptualization, Investigation, Writing - review & editing. **M. Salto-Tellez:** Conceptualization, Funding acquisition, Investigation, Methodology, Pro-

ject administration, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Nowogrodzki A. How to support open-source software and stay sane. *Nature* 2019;571(7763):133–5.
- [2] Marée R. Open practices and resources for collaborative digital pathology. *Front Med (Lausanne)* 2019;6:255.
- [3] Pell RA-O, Oien K, Robinson M, Pitman H, Rajpoot N, Rittscher J, et al. The use of digital pathology and image analysis in clinical trials. (2056-4538 (Electronic)).
- [4] Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArd DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017;7(1):16878.
- [5] Cancer Research UK. (2016) Testing times to come? An evaluation of pathology capacity across the UK. Available from: http://www.cancerresearchuk.org/sites/default/files/testing_times_to_come_nov_16_cruc.pdf Last accessed: Oct 12th 2020.
- [6] Report of the Independent Cancer Taskforce, UK, (July 2015) Achieving world class cancer outcomes: A strategy for England 2015–2020. Available from: https://www.cancerresearchuk.org/sites/default/files/achieving_world-class_cancer_outcomes_-_a_strategy_for_england_2015-2020.pdf Last accessed: Oct 12th 2020.
- [7] Karakusevic S, Edwards N, Lewis R and Dayan M (2016) The future of pathology services. Nuffield Trust. Available from: <https://www.nuffieldtrust.org.uk/research/the-future-of-pathology-services> Last accessed Oct 12th 2020.
- [8] Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArd DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017;7(1):16878.
- [9] Humphries MP, McQuaid S, Craig SG, Bingham V, Maxwell P, Maurya M, et al. Critical appraisal of programmed death ligand 1 reflex diagnostic testing: current standards and future opportunities. *J Thorac Oncol.* 2019;14(1):45–53.
- [10] Craig SG, Humphries MP, Alderdice M, Bingham V, Richman SD, Loughrey MB, et al. Immune status is prognostic for poor survival in colorectal cancer patients and is associated with tumour hypoxia. *Br J Cancer* 2020.

- [11] Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. *Methods* 2014;70(1):59–73.
- [12] Studer L, Blank A, Bokhorst JM, Nagtegaal ID, Zlobec I, Lugli A, et al. Taking tumour budding to the next frontier - a post-ITBCC 2016 review. *Histopathology* 2020.
- [13] Abduljabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med* 2020;26(7):1054–62.
- [14] Tsakiroglou AM, Fergie M, Oguejiofor K, Linton K, Thomson D, Stern PL, et al. Spatial proximity between T and PD-L1 expressing cells as a prognostic biomarker for oropharyngeal squamous cell carcinoma. *Br J Cancer* 2020;122(4):539–44.
- [15] Feng Z, Bethmann D, Kappler M, Ballesteros-Merino C, Eckert A, Bell RB, et al. Multiparametric immune profiling in HPV-oral squamous cell cancer. *JCI Insight* 2017;2(14).
- [16] Humphries MP, Bingham V, Abdullahi Sidi F, Craig SG, McQuaid S, James J, et al. Improving the diagnostic accuracy of the PD-L1 test with image analysis and multiplex hybridization. *Cancers (Basel)* 2020;12(5).
- [17] Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* 2020.
- [18] Arends MJ, Salto-Tellez M. Low-contact and high-interconnectivity pathology (LC&H Path): post-COVID19-pandemic practice of pathology. *Histopathology* 2020;77(4):518–24.
- [19] Sevinc A. Web of science: a unique method of cited reference searching. *J Natl Med Assoc* 2004;96(7):980–3.
- [20] del Barco BI, Stephan-Otto Attolini C, Slobodnyuk K, Igea A, Gregorio S, Gawrzak S, et al. Regulation of mammary luminal cell fate and tumorigenesis by p38 α . *Stem Cell Rep* 2018;10(1):257–71.
- [21] Taurillo DVF, Palomo-Ponce S, Stork D, Berenguer-Llgero A, Badia-Ramentol J, Iglesias M, et al. TGF β drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* 2018;554(7693):538–43.
- [22] Momcilovic M, Jones A, Bailey ST, Waldmann CM, Li R, Lee JT, et al. In vivo imaging of mitochondrial membrane potential in non-small-cell lung cancer. *Nature* 2019;575(7782):380–4.
- [23] Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* 2020;38(5):586–99.
- [24] Wang Z, Wu VH, Allevato MM, Gilardi M, He Y, Luis Callejas-Valera J, et al. Syngeneic animal models of tobacco-associated oral cancer reveal the activity of in situ anti-CTLA-4. *Nat Commun* 2019;10(1):5546.
- [25] Friedman G, Levi-Galibov O, David E, Bornstein C, Giladi A, Dadiani M, et al. Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+ CAFs to clinical outcome. *Nature Cancer* 2020;1(7):692–708.
- [26] Pell R, Oien K, Robinson M, Pitman H, Rajpoot N, Rittscher J, et al. The use of digital pathology and image analysis in clinical trials. *J Pathol: Clin Res* 2019;5(2):81–90.
- [27] Reichling C, Taieb J, Derangere V, Klopfenstein Q, Le Malicot K, Gornet J-M, et al. Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage III colon cancer outcomes in PETACC08 study. *Gut* 2020;69(4):681.
- [28] Humphries MP, McQuaid S, Craig S, Bingham V, Maxwell P, Maurya M, et al. Critical appraisal of PD-L1 reflex diagnostic testing: current standards and future opportunities. *J Thorac Oncol* 2018;15:1556.
- [29] Acs BA-O, Pelekanou V, Bai Y, Martinez-Morilla S, Toki M, Leung SCY, et al. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. (1530-0307 (Electronic)).
- [30] Roberti MP, Yonekura S, Duong CPM, Picard M, Ferrere G, Tidjani Alou M, et al. Chemotherapy-induced ileal crypt apoptosis and the ileal microbiome shape immunosurveillance and prognosis of proximal colon cancer. *Nat Med* 2020;26(6):919–31.
- [31] Talbot J, Hahn P, Kroehling L, Nguyen H, Li D, Littman DR. Feeding-dependent VIP neuron-ILC3 circuit regulates the intestinal barrier. *Nature* 2020;579(7800):575–80.
- [32] Jäkel S, Agirre E, Mendanha Falcão A, van Bruggen D, Lee KW, Knuesel I, et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* 2019;566(7745):543–7.
- [33] Choudhury SR, Babes L, Rahn JJ, Ahn B-Y, Goring K-AR, King JC, et al. Dipeptidase-1 is an adhesion receptor for neutrophil recruitment in lungs and liver. *Cell*. 2019;178(5):1205–21.e17.
- [34] Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, et al. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Cell*. 2019;178(6):1493–508.e20.
- [35] Welz P-S, Zinna VM, Symeonidi A, Koronowski KB, Kinouchi K, Smith JG, et al. BMAL1-driven tissue clocks respond independently to light to maintain homeostasis. *Cell*. 2019;177(6):1436–47.e12.
- [36] Chen W-T, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, Corthout N, et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell*. 2020;182(4):976–91.e19.
- [37] Nejmán D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 2020;368(6494):973.
- [38] Persson EK, Verstraete K, Heyndrickx I, Gevaert E, Aegerter H, Percier J-M, et al. Protein crystallization promotes type 2 immunity and is reversible by antibody treatment. *Science* 2019;364(6442):eaaw4295.
- [39] Owens R, Gilmore E, Bingham V, Cardwell C, McBride H, McQuaid S, et al. Comparison of different anti-Ki67 antibody clones and hotspot sizes for assessing proliferative index and grading in pancreatic neuroendocrine tumours using manual and image analysis. *Histopathology* 2020.
- [40] Liu K, Xia W, Qiang M, Chen X, Liu J, Guo X, et al. Deep learning pathological features in endemic nasopharyngeal cancer: Prognostic value and potential role for individual induction chemotherapy. *Cancer Med* 2020;9(4):1298–306.
- [41] Cummings M, King H, Hurst J, Tanner G, Khazin L, Thompson P, et al. Decreasing formalin concentration improves quality of DNA extracted from formalin-fixed paraffin-embedded tissue specimens without compromising tissue morphology or immunohistochemical staining. *J Clin Pathol* 2020;73(8):514.
- [42] Rafei-Shamsabadi D, Lehr S, von Bubnoff D, Meiss F. Successful combination therapy of systemic checkpoint inhibitors and intralesional interleukin-2 in patients with metastatic melanoma with primary therapeutic resistance to checkpoint inhibitors alone. *Cancer Immunol Immunother* 2019;68(9):1417–28.
- [43] Bezemer RE, Schoots MH, Timmer A, Scherjon SA, Erwich JJHM, van Goor H, et al. Altered levels of decidual immune cell subsets in fetal growth restriction, stillbirth, and placental pathology. *Front Immunol* 2020;11:1898.
- [44] Sanghvi AB, Allen EZ, Callenberg KM, Pantanowitz L. Performance of an artificial intelligence algorithm for reporting urine cytopathology. *Cancer Cytopathol* 2019;127(10):658–66.
- [45] Martino F, Varricchio S, Russo D, Merolla F, Iardi G, Mascolo M, et al. A machine-learning approach for the assessment of the proliferative compartment of solid tumors on Hematoxylin-Eosin-stained sections. *Cancers* 2020;12(5):1344.
- [46] Coulson-Gilmer C, Humphries MP, Sundara Rajan S, Droop A, Jackson S, Condon A, et al. Stanniocalcin 2 expression is associated with a favourable outcome in male breast cancer. *J Pathol: Clin Res* 2018;4(4):241–9.
- [47] Gray RT, Loughrey MB, Bankhead P, Cardwell CR, McQuaid S, O'Neill RF, et al. Statin use, candidate mevalonate pathway biomarkers, and colon cancer survival in a population-based cohort study. *Br J Cancer* 2017;116(12):1652–9.
- [48] Gray RT, Cantwell MM, Coleman HG, Loughrey MB, Bankhead P, McQuaid S, et al. Evaluation of PTGS2 expression, PIK3CA mutation, aspirin use and colon cancer survival in a population-based cohort study. (2155-384X (Print)).
- [49] Norberg KJ, Liu X, Fernández Moro C, Strell C, Nania S, Blümel M, et al. A novel pancreatic tumour and stellate cell 3D co-culture spheroid model. *BMC Cancer* 2020;20:1–13.
- [50] Gray M, Turnbull AK, Ward C, Meehan J, Martínez-Pérez C, Bonello M, et al. Development and characterisation of acquired radioresistant breast cancer cell lines. *Radiat Oncol* 2019;14(1):64.
- [51] Viratham Pulsawatdi A, Craig SG, Bingham V, McCombe K, Humphries MP, Senevirathne S, et al. A robust multiplex immunofluorescence and digital pathology workflow for the characterisation of the tumour immune microenvironment. *Mol Oncol* 2020.
- [52] Bellamy COC, Prost S. Multiplex tissue imaging: An introduction to the scope and challenges. *Am J Transplant* 2020;20(4):915–7.
- [53] Vejar MR, LaHaye N, Linstead EJ, Kim CS. Mine tailings particles with more homogenous/internal arsenic distribution pose a greater long-term exposure risk than particles with primarily surface-bound arsenic. *Change* 12(15):41.
- [54] Humphries MP, Craig SG, Kacprzyk R, Fisher NC, Bingham V, McQuaid S, et al. The adaptive immune and immune checkpoint landscape of neoadjuvant treated esophageal adenocarcinoma using digital pathology quantitation. *BMC Cancer* 2020;20(1):500.
- [55] Loughrey MB, Bankhead P, Coleman HG, Hagan RS, Craig S, McCorry AMB, et al. Validation of the systematic scoring of immunohistochemically-stained tumour tissue microarrays using QuPath digital image analysis. *Histopathology* 2018.
- [56] Bankhead P, Fernandez JA, McArt DG, Boyle DP, Li G, Loughrey MB, et al. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. *Lab Invest* 2018;98(1):15–26.
- [57] Kather JN, Suarez-Carmona M, Charoentong P, Weis C-A, Hirsch D, Bankhead P, et al. Topography of cancer-associated immune cells in human solid tumors. *eLife* 2018;7:e36967.
- [58] Humphries MP, Hynes S, Bingham V, Cougot D, James J, Patel-Socha F, et al. Automated tumour recognition and digital pathology scoring unravels new role for PD-L1 in predicting good outcome in ER-/HER2+ breast cancer. *J Oncol* 2018;2018:2937012.
- [59] Zheng H, Baranova K, Song J, Yan L, Biswas S, Chakrabarti S, et al. Overexpression of long noncoding RNA HOTAIR is a unique epigenetic characteristic of myxopapillary ependymoma. *J Neuropathol Exp Neurol* 2020.
- [60] Kane GM, Grünwald BT, Jang G-H, Masoomian M, Picardo S, Grant RC, et al. GATA6 expression distinguishes classical and basal-like subtypes in advanced pancreatic cancer. *Clin Cancer Res* 2020;26(18):4901.
- [61] Abengozar-Muela M, Esparza MV, Garcia-Ros D, Vázquez CE, Echeveste JI, Idoate MA, et al. Diverse immune environments in human lung tuberculosis granulomas assessed by quantitative multiplexed immunofluorescence. *Mod Pathol* 2020.

- [62] Echle A, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hutchins GGA, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology* 2020.
- [63] Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun* 2020;11(1):3877.
- [64] Kulkarni PM, Robinson EJ, Sarin Pradhan J, Gartrell-Corrado RD, Rohr BR, Trager MH, et al. Deep learning based on standard H&E images of primary melanoma tumors identifies patients at risk for visceral recurrence and death. *Clin Cancer Res* 2020;26(5):1126.
- [65] Sha L, Osinski B, Ho I, Tan T, Willis C, Weiss H, et al. Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. *J Pathol Inf* 2019;10(1):24.
- [66] Sha L, Osinski BL, Ho IY, Tan TL, Willis C, Weiss H, et al. Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. *J Pathol Inf* 2019;10:24.
- [67] Woerl A-C, Eckstein M, Geiger J, Wagner DC, Daher T, Stenzel P, et al. Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides. *Eur Urol* 2020;78(2):256–64.
- [68] Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, Darzi A, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020;26(9):1351–63.
- [69] Ruslan Kupriev DP, Paweł Redzyński, Casper da Costa-Luis, Saugat Pachhai, Alexander Schepanovski, et al. DVC: data version control - git for data & models. 2021.
- [70] Saranti A, Taraghi B, Ebner M, Holzinger A, editors. Property-based testing for parameter learning of probabilistic graphical models. *Machine learning and knowledge extraction*; 2020. Cham: Springer International Publishing; 2020.
- [71] Kargl M, Regitnig P, Müller H, Holzinger A. Towards a better understanding of the workflows: modeling pathology processes in view of future AI integration. In: Holzinger A, Goebel R, Mengel M, Müller H, editors. *Artificial intelligence and machine learning for digital pathology: state-of-the-art and future challenges*. Cham: Springer International Publishing; 2020. p. 102–17.
- [72] Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence—the third revolution in pathology. *Histopathology* 2019;74(3):372–6.
- [73] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Min Knowl Disc* 2019;9(4):e1312.