

# Will Digital Pathology be as Disruptive as Genomics?

Steven N. Hart<sup>1</sup>

<sup>1</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

Received: 09 April 2018

Accepted: 20 June 2018

Published: 19 July 2018

## Abstract

Digital pathology is the science of performing traditional pathological assessment in a digital environment. A digital transition is long overdue since histochemical analysis such as hematoxylin and eosin staining has remained unchanged in over 100 years. Importantly, the digitization of whole slide images further lends itself to advances in computational pathology and artificial intelligence to transform qualitative assessment into quantitative assessment. The impact of this transition from a computational infrastructure perspective is reminiscent of a similar transition in the field of genomics. In this article, I describe some of the similarities between genomics and digital pathology as well as highlight some key lessons learned to prevent the same mistakes and delays that slowed the genomics revolution.

**Keywords:** Genomics, bioinformatics, digital pathology

## INTRODUCTION

Without question, the field of genomics disrupted the way science is done. Innovative methods to sequence DNA at a relatively low cost and high throughput made previously cost-prohibitive studies, such as whole genome sequencing, almost commonplace. Genetic epidemiology migrated from chip-based microarray methods that served to estimate genome diversity to measuring it at base-pair resolution. This led to more fine mapping that focused on identifying the causative rare variants as opposed to surrogate markers that happened to be in linkage disequilibrium with a common variant. This level of resolution, accuracy, and value from sequencing technologies has made sequencing the *de facto* standard in both research and clinical contexts. Discoveries are now commonly made by sequencing the entire genome of patients, rather than only a handful of candidate genes. In <2 decades, the cost of a whole genome has dropped from \$3B to nearly \$1000. Innovation and scale continue to make sequencing attractive for the foreseeable future, particularly in the field of molecular pathology. At the Mayo Clinic alone, about 50 new genomics tests are developed annually.

Excitingly, advances in technology in the digital era are also driving the field of digital pathology forward at a pace to rival the genomics boom. Whereas genomics disrupted the status quo for genetic sequencing, digital pathology will be just as paradigm shifting (if not more so). For the most part, pathology has not

significantly changed in 100 years. Diagnoses are made by highly trained medical experts through qualitative or semi-quantitative descriptions on stained tissue sections mounted on glass slides and observed under light microscopy. Looking forward, many have recognized the need to move to a digital infrastructure.<sup>[1-3]</sup> This emerging field of digital pathology has shown tremendous benefits for organization, analysis, sharing, teaching, telepathology, quantitative and reproducible diagnoses, and others.

Regulatory challenges have limited the adoption in the United States but have been successfully overcome in Canada and Europe. However, in April 2017, in a landmark achievement, the Philips IntelliSite whole slide imaging system was granted the Food and Drug Administration (FDA) approval for the review and interpretation of digital surgical pathology slides.<sup>[4]</sup> In this heroic effort, Mukhopadhyay *et al.* conducted 2000 surgical pathology cases using tissue from multiple anatomic sites to show the safety and efficacy of a whole slide imaging (WSI) platform which was equivalent to light microscopy.<sup>[5]</sup> Now that the threshold for comparison

**Address for correspondence:** Dr. Steven N. Hart,  
Department of Health Sciences Research, Mayo Clinic,  
Rochester, Minnesota, USA.  
E-mail: hart.steven@mayo.edu

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

**How to cite this article:** Hart SN. Will digital pathology be as disruptive as genomics?. *J Pathol Inform* 2018;9:27.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2018/9/1/27/237143>.

### Access this article online

#### Quick Response Code:



**Website:**  
[www.jpathinformatics.org](http://www.jpathinformatics.org)

**DOI:**  
10.4103/jpi.jpi\_25\_18

(a.k.a. predicate devices) has been established, the FDA has now classified WSI systems as a Class II medical device. This somewhat simplifies the process for comparable devices to be approved. It is this classification that will spur significant investment by academia and industry to adopt, expand, and innovate in a digital pathology environment. This marks the beginning of an arms race in digital pathology to compete for the 3–8 billion dollar market.<sup>[6]</sup>

So why compare and contrast digital pathology and genomics? Most comparisons for the digitization of pathology are made to radiology,<sup>[7]</sup> where digitization was necessary given the nature of the data. Here, we posit that genomics is also a suitable corollary for digital pathology due to similarities in data size, analytical complexity, and the disruptive effect it will have in pathology. The size of data from digital pathology such as WSI is actually more comparable to genomics than radiology. WSI are typically 1–4 GB per slide, compared to 0.08 GB for X-ray images, 0.1 GB for MR, and 0.5 for computed tomography scans<sup>[8]</sup> compared to the typical 2 GB human genome variant file. The 10–100-fold difference in data size can overwhelm existing compute and storage infrastructures, thus warranting as significant investment in computational infrastructure – much like what was observed in the genomics revolution. This infrastructure is necessary since, like genomics, analysis of a single hematoxylin and eosin (H&E) may require multiple computational assessments. For example, a whole genome sequence analysis may require SNV and indel detection, copy number variation, structural variation calling, microsatellite analysis, and various measures of quality control, followed by manual inspections of regions of interest. H&E analytics for WSI in breast cancer tissues may require algorithms for detecting nuclei, quantitative the degree of pleomorphism, mitosis detection, classifying lobular involution, counting infiltrating immune cells, etc. Finally, the explosion of genomics has led to a recent and ongoing struggle to interpret, manage, share, and exploit value for such clinical testing paradigms. The same can be expected for pathology as the field becomes a more quantitative than qualitative discipline.

As the field of bioinformatics has grown alongside genomics, many of the same professionals that were responsible for implementing genomics will be leading efforts to operationalize digital pathology. Many of the lessons learned from genomics can, therefore, be applied.

## THE COST OF DATA

The significant expense can be invested to generate digital data from genomics and digital pathology. This section describes costs associated with these methods, comparing and contrasting the similarities and differences.

Cost is an important consideration for adopting new technologies: perhaps less so for major academic institutions, but smaller departments with limited capital much chose carefully about how they invest their limited resources. There are many definitions of cost, including acquisition

cost (capital cost of purchasing the instrument), an operating cost (how much money is spent every time the instrument is run), interpretation cost (e.g., what analytics, people, processes have to be in place to understand the data from the instrument), storage costs, and opportunity cost (gains and losses derived from an instrument given a fixed set of resources). Each of these costs is described below.

### Capital cost

The first – and perhaps the easiest to understand for administrators who generally approve large-scale investments and strategies – is the capital cost of the instrumentation or the acquisition cost. DNA sequencers can have different costs according to scale. For instance, Illumina offers a low throughput sequencer, MiSeq, that produces 8 gigabytes (Gb) in a 24-h run and a high throughput version, NovaSeq, produces 6 terabytes (Tb) in a 44 h run.<sup>[9]</sup> These instruments are roughly \$99K and \$985K, respectively, equating to \$34/Gb/day for the MiSeq and \$0.90/Gb/day for the NovaSeq. A parallel for digital pathology would be slide scanners. A low throughput slide scanner would be akin to a \$25K DigiPath Motic EasyScan Pro that can digitize 80 slides per day, whereas a high throughput alternative would be a Leica Aperio AT2 (700 slides per day, \$200K). Normalized by the amount of data generated each day, the cost of slide scanners are significantly lower than for sequencers, with low throughput being \$0.86/Gb/day and high throughput being upwards of \$0.78/Gb/day (assuming a 1 Gb file size). However, the sequencers have the added advantage of being able to sequence a single sample or many thousands of samples simultaneously, depending on the application. Barring tissue microarray, most slides correspond to a single patient and single stain. Note these estimates only relate to the capital expense and no additional processing that needs to be performed on a per sample basis, which is described below.

### Data acquisition cost

Aside from the instrumentation, the cost to generate data is more favorable for digital pathology. Sequencers have a fixed amount of sequencing capacity that will always be used for every run. The reagent cost per run is fixed, so it does not matter if there is 1 sample or many. In order to make DNA amenable to sequencing, complex molecular biology is required over a period of 3–5 days (depending on the protocol). Given the ultra-high sensitivity of sequencing instruments, the steps may have to be performed in one or more different physically isolated rooms so as not to introduce contamination. Every time a sequencer is run, it cost the operator about \$1000 in reagents for a MiSeq and \$9000 for a NovaSeq. Conversely, generating H and E is trivial, requiring <5 h and <\$5 to generate. However, diagnostic H and E is the standard of care for many pathology laboratories, so this cost is essentially free. No additional building spaces are needed, and the slide processing protocol is identical. This gives digital pathology a tremendous fiduciary advantage as a data generation device.

### Analytical cost

One often overlooked cost is the analysis costs. Bioinformaticians, those who traditionally lead the technical

analysis of these applications, are basically data scientists with specialized domain knowledge in biology. Sequencers generate billions of reads that require extensive quality control, alignment to a reference genome, identification of variants, functional annotation of variants, and removing false positives, all of which are performed by a bioinformatician. While some commercial graphical user interfaces (GUI) do exist and claim to perform this analysis, uptake by the community has been limited. It is critical to have a team capable of explaining why a particular algorithm made a particular decision. As of today, digital pathology requires less bioinformatics support, but this is rapidly changing and is a principal reason to expect widespread adoption of digital pathology. Tools for visualizing WSI and annotation of regions of interest are fairly simple to use, but the real power of digital pathology is the advances made in artificial intelligence or perhaps more appropriately augmented human intelligence (AHI). Algorithms are capable of detecting rare events such as mitoses, localizing metastatic breast cancer in lymph nodes, and classifying skin cancer with expert level accuracy. It is the bioinformaticians who will be applying new types of AHI algorithms to clinical use cases and automating repetitive tasks to increase the efficiency of the laboratory. All digital pathology applications should consider adding bioinformaticians to their teams' full-time. While increasing the total cost of analysis, their contribution will be invaluable as these advanced techniques evolve.

### Storage cost

There are also costs associated with storing and accessing data. As stated above, sequencer can generate 8–3000 Gb of data per day. Assuming 1 Gb for WSI file size, a fully operational WSI scanner would then generate between 80 and 700 Gb per day. Given the same capital investment for a sequencing instrument, 5–50 scanners could be purchased, leading to ~4,000 Gb per day, which is more data than the largest sequencer. This is not to say that the services these two platforms provide are interchangeable and rather establish an understanding between the impact of spending a fixed amount of capital on a particular problem and how that could affect data storage capacity. Storage is relatively inexpensive (\$0.01–0.02/GB/month) but becomes a nontrivial expense as data are generated over time. At this scale, it is important to consider the need for Information Technology (IT) support as these experts are responsible for ensuring adequate disk space, data security, planning for future investments, and strategic cost reductions. Even if digital pathology adopts a “cloud-first” data management strategy, there are different tiers of data storage, and this support will be necessary for maintaining interoperability with the cloud providers. It is the IT professionals who will be required to adhere to regulatory guidelines, institutional policies, and institutional best practices for data management.

### Opportunity cost

Separately, one has to consider not only cost but also value. Data are more valuable than gold: the more you use it and the

more you have of it, the more valuable it becomes. This is especially the case in the age of artificial intelligence. Highly curated training data are essential for discovery and validation, and these high-quality datasets are necessary to develop, test, and refine the computational algorithms that will drive the field of digital pathology. Those who recognize the intrinsic value of data will be more likely to lead the innovation and practice implementations. Therefore, the opportunity cost of not implementing and innovating a digital pathology is high.

## LESSONS LEARNED FROM GENOMICS THAT INFORM DIGITAL PATHOLOGY GROWTH

There are at least five significant “lessons learned” during the growth of genomics that are well suited to be transferred to digital pathology. Below, concrete examples provide a roadmap for success if realized sooner rather than later.

### Partnership, teamwork, and innovation

In genomics, it did not take long for users to generate data a faster rate than they could interpret it. Scientists who were early adopters quickly realized that the data did not fit into Excel spreadsheets. Laboratories did not typically have the command-line skills or computational and mathematical backgrounds to understand the data. Out of necessity, a cadre of bioinformaticians evolved from either the computer science or biology discipline. These specialists were in high demand to develop custom algorithms so that investigators could ask biologically relevant questions about the data, manage the growth of data, convert from one file format to another, etc. This led to the tongue in cheek label of “data janitors”<sup>[10]</sup> and the myth of “push-button bioinformatics.”<sup>[11]</sup> While it may seem innocent, these stigmas have been a detriment to the informatics community. Informaticians became viewed as support staff rather than collaborators even though they may be responsible for increasing the value of the data. Beyond the simple translation of mapping reads and calling small genetic variations, bioinformaticians built algorithms to detect microsatellite instability, copy number variations, translocations, inversions, as well as new ways to visualize and integrate data. Bioinformaticians bring unique expertise to understanding how computation can be used to solve clinically relevant problems and should be treated as equal collaborators and not just “data janitors.” Now, more than ever, the contributions made by informatics professionals are being given equal weight in decision-making, formulating and writing research grants, and directing large research programs.

Digital pathology will benefit more from bioinformaticians if their value is appreciated at a faster pace than it was in genomics. After all, bioinformaticians will be the ones building new slide viewers, new algorithms to count nuclei, and integrating imaging and nonimaging data for decision support. The sooner their value is recognized, the sooner more innovation will occur.

## Competition spurs innovation

One way to spur innovations is through the organization of “challenges” or “code-a-thons.” By providing access to labeled training (and unlabeled test data), challenge organizers encourage participation through either monetary compensation or “bragging rights” in a particular domain. In genomics, a multinational collaboration was undertaken to produce a ground truth set that was previously unthinkable. The genome in a bottle consortium,<sup>[12]</sup> led by the National Institute of Standards and Technology, extensively characterized a commonly used set of DNA samples as well as established a reference material to which all manners of genome sequencing analyses could be compared. By establishing a ground truth for reference materials that were available to the community, a framework was established for comparing accuracy of sequencing modalities and informatics methods.<sup>[13]</sup> This work was later extended by the FDA as part of their Precision FDA Challenge series.<sup>[14]</sup> Submitters and their scores are publicly displayed relative to the evaluation data, allowing competitors and collaborators to compare accuracy of products side-by-side for a particular task.

Similar competitions have also been held for digital pathology. Exemplary examples of successful competitions include the Assessment of Mitosis Detection Algorithms,<sup>[15]</sup> Tumor Proliferation Assessment Challenge,<sup>[16]</sup> and CAMELYON challenges.<sup>[17]</sup> The CAMELYON challenge was an especially interesting competition whose aim was to identify all pixels containing metastatic breast cancer from a H and E-stained lymph node WSI. The results of CAMELYON showed the accuracy of the top 5 algorithms were as accurate as an expert pathologist, and many were better than non-specialist pathologists. This is an exciting demonstration of the potential of AHI at democratizing geographically local expertise since the algorithm could be deployed on thousands of computers across the world-bringing standardized, reproducible, and highly accurate diagnoses to the masses.

## Transparency and reproducibility

While qualitative subjective assessments in a digital environment are a profound move forward in the right direction, the real innovation in digital pathology will come from advances in AHI. These algorithms will be developed over time using the expert knowledge of disease-specific pathologists. That said, the transition to quantitative assessment will be difficult. The expectation of AHI algorithms to improve diagnostic accuracy and throughput will be conditioned on the interpretability of the outputs. Luckily, this is an area of active research in the entire machine learning field, not just digital pathology. One such explainability model is Local Interpretable Model-Agnostic Explanation.<sup>[18]</sup> Importantly, these explanations are not a “nice-to-have” but may be required from a regulatory standpoint.<sup>[19]</sup> It is imperative to point out that AHI algorithms are meant to help guide pathologists – not replace them. They represent another tool at the pathologist’s disposal to be used in the appropriate context with the appropriate amount of weight in their interpretation.

Another aspect to transparency is open-source algorithms. Often, the ability to trust an algorithm is dependent on access to minute details that were used to make a decision or inference. An example in genomics would be whether or not duplicate reads were removed or flagged as this can significantly affect variant calling algorithms. Without computational experts to view the code, it becomes a black box and therefore cannot be improved for the use cases at hand. Some commercial genomics companies have managed to exist with closed source applications but struggle since it is highly challenging to keep up with the pace of discovery. For an example, the Genome Analysis ToolKit (GATK) started out as an open-source project, gaining contributors and knowledge from the community of developers in genomics, and enforcing the community standards for input and output file formats. However, in 2015, the Broad Institute opted for a commercial license model,<sup>[20]</sup> but a swift negative reaction from the community forced a reversal of this decision<sup>[20,21]</sup> and is now covered by a more permissive BSD license. In a completely closed system, a bioinformatician would be unable to explore and modify the codebase – a common practice with open-source software when faced with edge cases for a particular assay or laboratory process. In the digital pathology realm, this would be akin to the OpenSlide library<sup>[22]</sup> going to closed source. OpenSlide is the entry point for a number of image-processing pipelines that normalize proprietary vendor image formats into a unified application programming interface (API) – a major boon for interoperability. As an open-source library, bioinformaticians and programmers can more easily debug errors in code or add new features as the need arises. Finally, OpenSlide sets the expectation for future libraries that they be open as well, which helps not only transparency but also interoperability.

## Focus on interoperability

Both genomics and digital pathology are rapidly evolving ecosystems. This places a considerable burden on infrastructure development teams who are responsible for supporting applications, adding new functionality, fixing bugs, etc. The problem can be compounded if the right architectural designs are not specified up front. Until recently, constructing genomic analysis pipelines usually meant building a monolithic codebase with multiple configuration options, temporary files, and output formats. These fragile architectures were usually built by bioinformaticians who were not necessarily formally trained in software design and good coding practices. As the industry has circled around standard specifications such as the VCF, BAM, and CRAM,<sup>[23]</sup> it has now become apparent that interoperability was underappreciated in the beginning. As new tools are developed or updated versions become available, informaticians are keen to put these updates in production. Swapping out algorithms has been made significantly easier through the use of formal pipeline languages such as WDL<sup>[24]</sup> and CWL.<sup>[25]</sup> Moreover, sharing data across institutions introduced other layers of complexity, which spawned the formation of the Global Alliance for Genomic Health (GA4GH, <https://www.ga4gh.org/>). Significant amounts of work have

been centrally coordinated to standardize APIs for information exchange to facilitate modularity and interoperability.

These concepts of modularity in coding architecture and enforced adoption of industry standards could benefit more rapid adoption of and progress in digital pathology. OpenSlide is only a minor piece of this intricate puzzle. Like VCF, BAM, and CRAM, a standard exists for image formats: Digital Imaging and Communications in Medicine. In fact, a recent connect-a-thon seemed to solidify the industry's resolve to adopt this standard<sup>[26]</sup> although changes will need to be made to OpenSlide to support this transition. Beyond reading images, standard APIs will also be needed for other aspects of digital pathology. For example, there are a multitude of whole slide annotation and visualization tools available. Some will be desirable for certain use cases but not others. What would be ideal is if developers would build more libraries (like OpenSlide) than full applications. This would ensure interoperability across platforms from slide viewing and annotating to APLIS integration and decision support.

### Positive user experience leads to adoption

There are two main types of users to software systems: those that operate on the command-line interface (CLI, e.g., bioinformaticians) and those that prefer a (GUI, e.g., consulting pathologists). Each user interface has its own challenges and demands, but a focus on the user experience is paramount to a software platform to be adopted using either interface.

In genomics, the first pieces of software developed were restricted to a CLI. It allowed programmers the flexibility to automate analyses and alter configuration parameters with ease. However, “help flags” may not contain sufficient information about how and where such parameters were used or how they may affect the final output. As the field evolved, CLI users required more extensive documentation – leading the adoption of entire websites (e.g., GATK) or standard documentation libraries (e.g., Sphinx<sup>[27]</sup> and ReadTheDocs<sup>[28]</sup>). Increased quality documentation allows users to have a better understanding for the application and leads to more adoption by the community.<sup>[29]</sup> At this stage, the digital pathology CLI landscape is underdeveloped. Few libraries are used across multiple projects, likely due to the lack of common standard formats. CLIs will likely play an integral part in adopting AHI algorithms, so harmonization is needed early in the process.

In digital pathology, the focus has been centered on developing GUIs. Development of GUIs is perhaps an even greater challenge than CLIs since it requires a visually appealing design as well as an ergonomic and functional user experience. While some nontraditional visualization paradigms are being developed (e.g., power walls<sup>[30]</sup> and virtual reality<sup>[31]</sup>), most GUIs are being developed for projection onto medical-grade monitors.<sup>[32]</sup> This represents a significant shift in operational requirements on the part of the pathologist, whose tactile interaction with the optical microscope is replaced by the more traditional keyboard, mouse, and monitor. It will be imperative to co-develop user-friendly interfaces with pathologists so that

their specific use cases can be met to minimize the discomfort of migrating to the digital environment.

There is, however, an intersection between CLI and GUI that has been quite successful in genomics: the Galaxy framework.<sup>[33]</sup> The aims of the Galaxy toolkit are to (1) make computationally complex analytics available to investigators with limited computational expertise, (2) construct complex, customizable, and reproducible workflows, and (3) publish those analysis to the web.<sup>[34]</sup> Importantly, the framework allows investigators access to all the options available through the CLI while abstracting away the system-level administration. The Galaxy framework is currently being used in the “Cloud-based Image Analysis and Processing Toolbox” project<sup>[35]</sup> but can easily be extended for additional digital pathology needs.

## DISCUSSION

Throughout this article, I have tried to highlight some of the similarities between digital pathology and genomics. However, there are key differences that were not discussed. For example, the clinical impact of genomics has proven its worth by becoming a dominant method in molecular pathology. The same cannot be said for digital pathology. Due to its relative immaturity, much of the work in digital pathology has been driven to satisfy the regulatory requirement of demonstrating equivalence to traditional light microscopy, rather than demonstrating improved clinical outcomes. Genomic technology too was once in infancy, and many early papers were essential to convince the field of its equivalency to orthogonal technologies. While not perfect, the corollary between genomics and digital pathology highlights significant opportunities for cross-disciplinary method development and a chance to benefit from some of the missteps and delays from the past.

Given the significantly lower capital investment, less intrusive nature, higher quality assertions, quantitative and reproducible advantages, and sample acquisition costs, digital pathology is expected to be as disruptive and potentially more so than genomics. These are not the only issues that we will face as we cross the digital divide, but they do enable a starting point for discussions regarding how the community will flourish.

### Acknowledgments

The author would like to thank the Mayo Clinic Center for Individualized Medicine and Department of Laboratory Medicine and Pathology for funding this work.

### Financial support and sponsorship

Nil.

### Conflicts of interest

There are no conflicts of interest.

## REFERENCES

1. Lloyd MC, Monaco JP, and Bui MM. Image Analysis in Surgical Pathology. Available from: <http://www.ncbi.nlm.nih.gov/>

- pubmed/27241112. [Last accessed on 2018 Mar 30].
2. Bueno G, Fernández-Carrobles MM, Deniz O, García-Rojo M. New trends of emerging technologies in digital pathology. *Pathobiology* 2016;83:61-9.
  3. Hanna MG, Pantanowitz L, and Evans AJ. Overview of Contemporary Guidelines in Digital Pathology: What is Available in 2015 and what Still Needs to be Addressed? Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25979986>. [Last accessed on 2018 Mar 30].
  4. Commissioner O of T. Press Announcements – FDA Allows Marketing of First Whole Slide Imaging System for Digital Pathology. Office of the Commissioner; Available from: <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm552742.htm>. [Last accessed on 2018 Mar 30].
  5. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, *et al*. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A Multicenter blinded randomized noninferiority study of 1992 cases (Pivotal study). *Am J Surg Pathol* 2018;42:39-52.
  6. Formats\*: Digital Pathology Market by Product (Whole Slide Imaging, Image Analysis Informatics, Information Management System Storage & Communication, Digital IVD Devices, and Others [Telepathology]), Type (Automated Hematology Analyzers & Blood Cell Counters, Chromosome Analyzers, FISH Enumeration Systems, Urine Sediment Analyzers, Immunohistochemistry Image Analysis Applications, Digital Cytopathology, Digital PCR, Signet Cell Detection, and Digital Holographic Microscopy), and End User (Educational, Clinical, and Pharma and Biotech Companies) – Global Opportunity Analysis and Industry Forecast, 2017-2023. In: Allied Market Research. Available from: <http://www.alliedmarketresearch.com/digital-pathology-market>. [Last accessed on 2018 Mar 30].
  7. Chetty R. Pathology and radiology taking medical ‘hermeneutics’ to the next level? *J Clin Pathol* 2017;70:553-4.
  8. Softnet. DICOM Library – About DICOM most Common Features of Study. Available from: <http://www.dicomlibrary.com/dicom/study-structure/>. [Last accessed on 2018 Mar 29].
  9. Sequencing and Microarray Systems. Available from: <http://www.illumina.com/systems.html>. [Last accessed on 2018 Mar 30].
  10. Lohr S. For Big-Data Scientists, “Janitor Work” is Key Hurdle to Insights. *NY Times*; 2014. Available from: <http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html>. [Last accessed on 2018 Mar 30].
  11. Smith DR. The battle for user-friendly bioinformatics. *Front Genet* 2013;4:187.
  12. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, *et al*. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol* 2014;32:246-51.
  13. Highnam G, Wang JJ, Kusler D, Zook J, Vijayan V, Leibovich N, *et al*. An analytical framework for optimizing variant discovery from personal genomes. *Nat Commun* 2015;6:6275.
  14. Precision FDA Challenges – Precision FDA. Available from: <http://www.precision.fda.gov/challenges>. [Last accessed on 2018 Mar 28].
  15. Veta M, van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, *et al*. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal* 2015;20:237-48.
  16. In: MICCAI Grand Challenge: Tumor Proliferation Assessment Challenge (TUPAC16); 23 September, 2013. Available from: <http://www.tupac.tue-image.nl/node/62>. [Last accessed on 2018 Mar 28].
  17. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199-210.
  18. Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?” Explaining the Predictions of Any Classifier; 2016. Available from: <http://www.arxiv.org/abs/1602.04938>. [Last accessed on 2018 Mar 30].
  19. Goodman B, Flaxman S. European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”; 2016. Available from: <http://www.org/abs/1606.08813>. [Last accessed on 2018 Mar 30].
  20. GATK Licensing Through Broad Starts Today. In: GATK-Forum. Available from: <https://www.gatkforums.broadinstitute.org/gatk/discussion/5458/gatk-licensing-through-broad-starts-today>. [Last accessed on 2018 Mar 29].
  21. Pjacock. In: Storify. Available from: <http://www.storify.com/pjacock/gatk-v2-0-not-open-source>. [Last accessed on 2018 Mar 29].
  22. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J Pathol Inform* 2013;4:27.
  23. Samtools. In: GitHub. Available from: <https://www.github.com/samtools/hts-specs>. [Last accessed on 2018 Mar 29].
  24. In: GATK-Forum. Available from: <https://www.gatkforums.broadinstitute.org/gatk/discussion/7349/the-art-of-the-pipeline-introducing-cromwell-wd>. [Last accessed on 2018 Mar 29].
  25. In: GitHub. Available from: <https://www.github.com/common-workflow-language/common-workflow-language>. [Last accessed on 2018 Mar 29].
  26. Clunie D, Hosseinzadeh D, Wintell M, De Mena D, Lajara N, Garcia-Rojo M, *et al*. Digital imaging and communications in medicine whole slide imaging connectathon at digital pathology association pathology visions 2017. *J Pathol Inform* 2018;9:6.
  27. Overview – Sphinx 1.8.0+Documentation. Available from: <http://www.sphinx-doc.org/en/master/>. [Last accessed on 2018 Jun 13].
  28. Home | Read the Docs. Available from: <https://www.readthedocs.org/>. [Last accessed on 2018 Jun 13].
  29. Karimzadeh M, Hoffman MM. Top considerations for creating bioinformatics software documentation. *Brief Bioinform* 2017. pii: bbw134.
  30. Treanor D, Jordan-Owers N, Hodrien J, Wood J, Quirke P, Ruddle RA, *et al*. Virtual reality powerwall versus conventional microscope for viewing pathology slides: An experimental comparison. *Histopathology* 2009;55:294-300.
  31. Farahani N, Post R, Duboy J, Ahmed I, Kolowitz BJ, Krinchai T, *et al*. Exploring virtual reality technology and the oculus rift for the examination of digital pathology slides. *J Pathol Inform* 2016;7:22.
  32. Fetterly KA, Blume HR, Flynn MJ, Samei E. Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. *J Digit Imaging* 2008;21:193-207.
  33. Schatz MC. The missing graphical user interface for genomics. *Genome Biol* 2010;11:128.
  34. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, *et al*. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* 2016;44:W3-10.
  35. Cloud-Based Image Analysis and Processing Toolbox. Available from: <http://www.cloudimaging.net.au/>. [Last accessed on 2018 Jun 13].