Scientific Discovery at the Press of a Button: Navigating Emerging Cloud Laboratory Technology

*D. Sebastian Arias Rebecca E. Taylor**

D. Sebastian Arias
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA

Prof. Rebecca E. Taylor
Department of Mechanical Engineering
Department of Biomedical Engineering
Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA
bex@andrew.cmu.edu

Keywords: *Cloud laboratory, automated research, remote research, self-driving labs, high-throughput techniques*

The "cloud lab," an automated laboratory that allows researchers to program and conduct physical experiments remotely, represents a paradigm shift in scientific practice. This shift from wet-lab research as a primarily manual enterprise to one more akin to programming bears incredible promise by democratizing a completely new level of automation and its advantages to the scientific community. Moreover, they provide a foundation on which automated science driven by artificial intelligence can be built upon and thereby resolve limitations in scope and accessibility that current systems face. With a focus on DNA nanotechnology, our group has had the opportunity to explore and apply the cloud lab to active research. In this perspective, we delve into the future potential of cloud labs in accelerating scientific research and broadening access to automation. We further explore the challenges associated with the technology in its current state, including difficulties in experimental troubleshooting, the limited applicability of its parallelization in an academic setting, as well as the potential reduction in experimental flexibility associated with the approach.

1 Introduction to the Cloud Lab

Over the past two decades, interest in automation technologies has rapidly increased (see **Figure 1**) due to the many benefits it can provide research.[1] However, current laboratory automation that is generally accessible to researchers comes in the form of specialized lab equipment that is typically limited to specific tasks. This inhibits the extent of the benefits that one can extract from these systems. As automation technologies have progressed, and with the rise of more advanced artificial intelligence (A.I.), the interest in A.I. driven research has similarly expanded. In fact, back in 2009, a "robot scientist", a system combining A.I. and automated research tools, became the first system to make a novel scientific discovery without human input.[2, 3, 4] More recently, "self-driving labs" (SDLs), systems that incorporate A.I. and advances in robotics and automation to vastly accelerate chemical space exploration, have rapidly garnered interest for accelerating the synthesis of materials in chemical and material sciences.[5] Although these systems have already proven useful in accelerating research, they are still heavily limited. In the analysis done by Abolhasani and Kumacheva, three crucial components that are lacking from current systems are: 1) standardization and cost-effective hardware, 2) readily accessible software, and 3) user-friendly operational guidelines.[5]

In parallel, there has been significant progress in the development of "cloud labs." These laboratories have been designed from the ground up with a focus on automation, serving as off-site research facilities which are remotely accessed by researchers. They aim to provide access to comprehensive and standardized automation of an expansive number of techniques fundamental to biological, chemical, and material science research and thereby completely remove scientists from the physical lab. In this new approach, researchers code experiments, which are performed at these remote laboratories. They then receive the generated results and can continue the usual iterative loop using this new workflow. This concept may

sound similar to those familiar with biofoundries, an approach that uses heavily automated, centralized facilities to streamline the design and manufacturing of biological systems.[6, 7] However, whereas biofoundries focus on automating specific segments of biomanufacturing processes, cloud labs aim to provide an entire laboratory environment suitable a wide range of scientific disciplines.[8]
The complete automation of experimental research targeted by cloud labs would be a significant leap within the domain of automated research beyond what is achieved through typical equipment-based approaches. Moreover, by serving as a foundation on which SDLs are built, they can help overcome their previously mentioned limitations. For one, they could provide standardized and well-documented software that is user-friendly and easily accessible. Appropriate cost-sharing funding approaches could also make them cost-effective options. Lastly and as a pivotal benefit, cloud labs that successfully automate a comprehensive range of techniques could broaden the applicability of self-driven labs by lifting the constraints on possible experimental approaches. Researchers would no longer need to develop the automated system and would instead need only focus on developing the software needed to implement their self-driven lab with a broad range of capabilities at their disposal. As a result, cloud labs could be the key to the rapid progression and proliferation of not only automated research workflows but also of SDLs across fields. In fact, researchers have already applied OpenAI's large language models to design an "agent" capable of guiding and performing research with a cloud lab serving as the automated laboratory.[9]
Our lab has been exploring this emerging technology as developed by the Emerald Cloud Lab (ECL). In their software interface experimental procedures are reduced to function calls. These functions can be strung together to program simple or complex experimental procedures that are run at their remote facility (see **Figure 2**). Behind the scenes, these experimental procedures are parsed into appropriate experimental instructions. The execution of the experiments, and execution queues, are managed by the software while taking into account various factors including availability of reagents and instruments and the nature and dependencies of the samples to optimize efficiency and maintain sample and protocol requirement. Their laboratory facilities use a hybrid automation model where human operators interface with automated instruments to complete protocols. This hybrid model broadens the number of available procedures by compensating for systems that do not have automated alternatives or those that are incompatible within the same network of systems. Once procedures are completed, generated data is automatically uploaded to the cloud for access and further analysis. With these capabilities researchers could set up remote and automated research workflows.
Fueled by this potential, there has already been substantial cloud lab adoption in the pharmaceutical industry. Academic interest still continues to increase, and notably, Carnegie Mellon University (CMU), in collaboration with the ECL, is building the first-ever academic cloud lab that is scheduled to become operational in the spring of 2024. In preparation for the opening of the CMU Cloud Lab, our group has began to transition the time-consuming and labor-intensive portions of our experimental research, namely the production and characterization of structural DNA nanotechnology, to the cloud. With an experience that may be representative of the broad user space in engineering, nanotechnology, and materials science, our lab has gained key insights into this developing technology, including its future potential and the considerable barriers in transitioning research to the cloud. Here we discuss these insights and share the lessons learned for effectively transitioning to a cloud lab workflow.

2 Promising Advantages of Cloud Labs

Although automation and its many benefits have already been previously discussed, we begin this section with a short overview while highlighting the key advantages that the cloud lab approach provides over traditional in-lab automation. This is followed by a discussion on the far-reaching benefits of cloud labs in broadening access and lowering barriers to scientific approaches as well as the implementation of advanced automated systems such as self-driven labs. We focus on the future potential of the technology as it reaches full maturity in this discussion, while the subsequent section will address current and future challenges.

2.1 Accelerating Science: Enhancing Pace and Quality

Automation is one of the core foundations of cloud labs. With this comes the usual benefits that have made it so appealing across different industries: parallelization and low downtime.[10, 11] Parallelization, the ability to complete processes using multiple concurrent threads, is achieved through strategic purchasing of high-throughput equipment alongside the more direct route of increasing the number of equipment available. When coupled with the capacity to run experimental procedures uninterrupted through all hours of the day and night, the potential for a rapid increase in the pace of scientific research and, as a result, scientific advancements is apparent. In stark contrast with typically implemented automation, fully developed cloud labs have the potential to completely automate laboratory workflows. This level of automation would not only extend the typical benefits but also affords the capability of remote research. Moreover, as previously highlighted, cloud laboratory research has the potential to increase reproducibility and directly address the reproducibility crisis in science.[8, 12] Issues with reproducibility not only undermine scientific credibility but also hinder progress by impeding the ability of other researchers to build upon and expand existing knowledge and to collaborate.[13, 14, 15] The implementation of typical automation can help. However, cloud labs, again, take these benefits a step further. This new paradigm could provide protocol standardization on a wide scale. Moreover, the automated generation of protocol documentation far exceeds that normally taken in a traditional lab as it would normally be too laborious to so.[13] This new standard for documentation could ensure comprehensive and consistent records, thereby enhancing reproducibility, data integrity, and facilitating collaboration among scientists.

Similarly, the standardization of wet-lab procedures, techniques, and approaches with built-in good research practices can facilitate reproducibility. Protocol standardization and sharing can minimize variations and discrepancies in experimental methods thereby promoting reliable replication of experiments. Jointly, these improvements lead to one of the most impactful consequence of cloud labs: the ease of sharing protocols. Within a cloud lab ecosystem, reproducing an experimental procedure is reduced to the sharing of code. With cloud labs, researchers could benefit from the same assumption that programmers today enjoy: consistent results regardless of the person running the program. Thus, the barrier for collaborating and building on previous results is vastly diminished. Large researcher-driven, open-source repositories and communities like Stack Overflow are sure to evolve driven by the growing demand for knowledge sharing and support. In fact, a GitHub community has already been set up for the Carnegie Mellon University Cloud Lab for this very purpose.[16, 12]

From the perspective of an active researcher and to the relief of graduate students across the world, the most tangible impact of cloud labs might be the liberation of the time bound to mundane tasks. Many aspects of scientific research can be repetitive and tedious. As a result, incredibly talented and highly educated individuals drudge through the exacting and repetitive tasks inherent to scientific research. Automating these tasks would free up researchers to focus on intellectual contributions to science such as experiment design, process development, and literature reviews. In other words, researchers could focus on the portions of science that drive innovation and foster creativity rather than manually mixing hundreds of samples. For a longer discussion on the benefits of automation in life science research, we direct the reader to the excellent article by Holland and Davies.[1]

2.2 Broadening Access and Lowering Barriers

Automation and emerging platform technologies have continued to give researchers new capabilities while software advances, such as machine learning, have reduced the complexity associated with instrument usage. For example, optical trap experiments once required custom-built hardware, but newer platforms, such as Lumicks, enable once-prohibitively challenging experiments to be run at the press of a button.[17] Similarly, the automation of atomic force microscopy studies has allowed for rapid acquisition and enhanced image quality previously achievable only by seasoned experts. Despite the incredible advances, the overall barrier of entry for researchers to apply a diverse set of research techniques remains considerably high. While these innovations have streamlined the utilization of particular experimental methodologies, their accessibility and advantages tend to be confined to isolated techniques or instrument-specific

platforms.
Likewise, the applicability of SDLs have been bounded to a relatively small subset of research topics and to an even smaller subset of the scientific community due to limitations in hardware and the prohibitively high expertise required.[5] Currently, such systems rely on custom-built hardware that is both expensive and that inherently restricts the variety and flexibility of research that can be conducted. In this reduced space, the science is dictated by the modules that are available. It is further limited by the difficulty in scaling up these systems, as adding modules to the workflow not only swells the cost but also the complexity and increases the points of failure. Due to their custom-built nature, the development and implementation of modules for an SDL necessitates significant expertise in robotics and specialized knowledge in hardware integration on top of the necessary expertise to conduct the research of interest. This presents an incredibly high barrier to entry that few can overcome.
In the envisioned cloud lab paradigm where complicated experimental procedures are broadly reduced to function calls, the learning curve associated with implementing new techniques is minimized. A researcher would no longer need to study the intricacies of performing an experiment, nor have a deep understanding of the equipment, nor need an abundant amount of practice to perfect it. This process can be reduced to reading the documentation of a function (or alternatively a quick visit to Cloud Lab Stack Overflow), letting the software determine optimal settings for samples, and letting the automated systems execute it. As in research today, a fundamental knowledge of techniques, their uses and capabilities would still be necessary, but applying them would be approachable in a way that we have not seen. Researchers could potentially skip the training loop required to gain experimental expertise altogether (See **Figure 3**). Of course, cloud labs present a learning curve. One must learn the language, be familiar with the intricacies of function calls, be aware of the capabilities and limitations, and be familiar with the difference in approach as compared to typical wet-lab research. The key difference is that hundreds of techniques lie over this curve, not just one.
It is this broad and accessible automation that positions cloud labs as a transformative solution to the challenges of current SDLs and systems alike. By providing a centralized platform with the flexibility to string together a diverse array of techniques to automate complex procedures, they could largely expand the possible scientific scope of SDLs. Moreover, they could eliminate the extensive expertise barrier as all hardware and software requirements are handled on the backend with no user interaction needed. As a result, the development and deployment of SDLs can be made approachable to a broad array of researchers across diverse fields.
This underpins another avenue by which cloud labs can increase access: the availability of instrumentation. Given the importance of platform technologies to scientific progress, it follows that lack of access to instrumentation is one of the key limitations of the traditional research model. Instrumentation can often be too expensive for laboratories with limited resources to acquire. Moreover, although grants can cover the initial cost of instruments, they rarely provide funds for the required maintenance and running costs. As a result, the potential benefit of an instrument for an individual research group might not justify the time or financial costs associated with it. Cloud labs could change the weights of this equation. With the technology, testing out a new procedure or technique would not necessitate acquiring new instrumentation nor vast expertise. As a result, one-off experiments on a range of instruments would be increasingly viable and quick offshoots that might lead to larger discoveries might occur more frequently. Science is only to benefit from such a large array of tools under researchers' fingertips.
Beyond the scientific bubble, cloud labs might serve to increase societal access to science. This drastic lowering of technical barriers could allow the broadening of scientific accessibility to underserved, underfunded institutions and populations that do not have the resources to attain all necessary materials and equipment needed to run sequences of experimental procedures. Moreover, it might allow the extension of these capabilities beyond a research focus to an educational one and thereby improve the efficacy of science education.[18, 19, 20] The capability of remote workflows could also remove geographical barriers, allowing individuals without local access to cutting-edge laboratories the opportunity to engage in advanced scientific research and education.

3 Current Challenges and Considerations for Future Users

Fully developed cloud labs that successfully automate the large majority of experimental techniques have immense potential. However, although the technology has made great strides, it has not reached this level of maturity. Early adopters of the cloud lab research model will face unique challenges. Some of these are expected to be transient and to improve, while others might be fundamentally incompatible with academic research and will need to be addressed if the technology is to become viable for broad usage in an academic setting.

3.1 Expanded Debugging Space

The cloud lab, as a fusion of software and hardware, faces the inherent challenges of both, and the task of seamlessly integrating their functionalities. The software must be able to communicate with hundreds of diverse types of instruments, track compatibility of thousands of objects, handle settings and configurations with dependencies, optimize the timings of experiments and instrument usage, and all seamlessly coalesce into an approachable and flexible interface that allows researchers to run series of experiments from the comfort of their office. Software bugs, as an unavoidable reality of software development, are present in the cloud lab. Unlike pure software, however, the cloud lab faces the challenges of experimental research brought by both hardware and humans: hardware can break and humans make mistakes. This interplay heavily expands the debugging space that typical researchers are accustomed to. Although it might be more familiar to those developing their own SDLs and systems alike.

Moreover, the remote element of cloud research is an aspect that we have found is surprisingly hard to get accustomed to. A normally overlooked yet keenly important set of skills that trained researchers develop throughout their career are the numerous checks that are performed while physically executing procedures. These are in most cases so habitual that they cease to register, yet they are a fundamental aspect of research. In traditional labs when experimental irregularities are encountered or suspected due to unexpected results, researchers carefully mark down those concerns, and they are considered when evaluating end results. If not encoded in the experimental procedure, these sorts of "quality checks" can be overlooked in automated procedures. Like many things, it is when you are unable to perform these checks that their importance becomes apparent.

Procedures and new approaches are therefore needed to compensate for the absence of live monitoring and ability to oversee experiments conducted in the cloud lab, such as the monitoring of pipetting pressure to determine liquid transfer success. When debugging an experiment "gone wrong" in a traditional lab, unexpected results can point to a variety of parameters that might need to be adjusted or corrected. The same procedure run at a cloud lab could point to those same parameters or to a range of other new cloud lab-specific issues. Understanding the cloud lab framework as an iceberg provides insights into potential failures. In this analogy the user interface is at the top. Protocols then progress deeper through the layers, through the back-end software and finally to laboratory operations, (see **Figure 4**). User-caused issues normally occur due to oversights that are generally caught at the user-interface level with useful error messages thrown. When they are not caught, they can continue down and cause issues at deeper levels, or result in a valid protocol with unwanted parameters. In either case, they are independently resolved through typical code debugging.

Issues with internal causes are far more difficult to approach, because they often cannot be completely diagnosed or resolved by the user. In these cases the cloud lab acts as a black box either due to restricted user access or the complexity of its intricate software, hindering users from gaining a complete understanding of its internal workings. At the user-interface level, these issues might prevent a researcher from uploading or verifying properly-formatted protocols. In the software back-end, they can cause parsing errors that result in an unwanted or invalid protocol. Finally, at the laboratory operations level, they might be related to an operator or a machine issue. To approach troubleshooting, users must determine the source, whether it was user related or internal and at what level their problem might have occurred. Partnering with internal troubleshooting teams is essential in the process.

Robustness of an automated system is crucial and underpins the reliability and consistency of operations. Cloud lab technology, in its current state, attempts to maintain a suitable level of robustness across its operations. The software is reasonably robust with few errors and bugs. Instrumentation is regularly calibrated, and users are able to freely review an instrument's last verification results. Additionally, human operators are trained on the usage of relevant instruments. Still, experimental procedures inherently present numerous potential points of failure, and the number of these points are greatly multiplied when automating the complex array of tasks required to fully automate a wet-lab environment. This amplification of failure points presents a formidable challenge in maintaining the system's robustness. Despite the progress that has been made in the technology, unresolved issues accumulate such that the system's consistency and robustness remain a key hurdle. The effectiveness of the system can be heavily dependent on use case. However, throughout our time navigating the technology and the adaptation and integration of our workflows to the cloud lab paradigm, we have encountered a spectrum of issues across the layers of the iceberg. These have ranged from minor and easily resolvable program bugs to critical errors that have resulted in the loss of valuable samples, due to both software issues and operator mishandling. **This underscores the critical necessity for further technological development as well as continual improvement to instrument management and upkeep and operator training if the cloud lab is to achieve an appropriate level of robustness.**

Moreover, although we have had access to such aid in transitioning research to the cloud lab, these issues could often not be diagnosed by us or internal teams. Many times we have been left with only one frustrating option to diagnose issues: repetition at a scale that would be unnecessary in traditional labs in hopes that one rerun might shed some light into whether the cause has a scientific source or whether it was due to an operator or machine error. Advancements in the form of newer instrumentation aimed at automation, improved software, and comprehensive training are expected to continue to address these issues. As the technology continues to develop, the tight controls, systematic and consistent labeling, and numerous other automatic checks performed by the cloud lab may end up reducing the overall frequency of errors when compared to traditional labs and could lead to the improved reproducibility that is one of the key potential benefits of the technology. Still, the issues currently present can inhibit researchers from transitioning to cloud-lab workflows without further advances in the technology. This underscores two further points. **First, cloud labs must substantially improve troubleshooting tools and error detection and tracking for this technology to be a viable alternative to physical labs for research in a broad number of fields. Second, prospective users need to understand that cloud labs introduce additional components that must be considered when interpreting results and prepare to engage in a new paradigm of troubleshooting approaches and tools. In part, this includes the implementation of comprehensive control experiments to accurately assess and ensure the integrity of the research findings within these complex systems.**

3.2 Narrow Efficacy of Parallelization

Parallelization is one of the most significant advantages of cloud labs with the promise that it could enable research acceleration in way akin to the transformative effect it has had in the computational space. It is clear how specific applications, such as generating experimental data to train large neural network models or examining a large chemical space, could significantly benefit from such acceleration. However, this is not widely beneficial to all academic research as a large portion of it comprises smaller-scale or tightly focused ventures.

Furthermore, in analyzing the efficacy of cloud labs for academic applications it is important to note that the parallelization that we can expect from cloud labs is not analogous to what we have seen in computational applications for one key reason: unlike computational hardware, which is applicable to any computational problem, experimental instruments are generally limited to one or a few specific functions. As a result, in a cloud lab, the level of parallelization is a case-by-case property. The acceleration one can expect for a specific experimental procedure is, therefore, governed by the type and number of instruments available for that specific purpose. Due to the instrumentation cost and financial incentives, the number of available instruments is governed by user demand. One may expect common procedures,

such as liquid handling, to reach significant levels of parallelization. However, less commonplace procedures might not have the demand to warrant more than a single instrument. In our experience, we have found that it is not uncommon for there to only be one instrument capable of doing a specific procedure. Moreover, the instruments that are available may lack the specific functionality that you might rely on. We experienced this firsthand when attempting to conduct agarose gel electrophoresis studies with custom loading buffers. Given the commonplace and essential nature of this procedure in our field we presumed that the available system would be capable of accommodating custom gel buffers. However, this was not the case and a limitation of the instrument itself.
To evaluate the efficacy of a cloud lab, researchers can consider several factors that will assist in the assessment process. **Determining the value of a cloud lab as an investment crucially depends on understanding the level of parallelization needed, verifying the availability of key capabilities, and setting realistic expectations for targeted procedures regarding the extent of parallelization achievable.** In the cases where parallelization is not essential, the ability of the cloud lab to automate processes might still provide great utility. For example, our lab aims to produce the majority of our DNA origami using the cloud lab to free up large portions of time. However, in the cases where there is no significant need for large-scale parallelization and where the cloud lab is only able to accomplish the more basic portions of research, one might question: can smaller and simpler alternatives, such as automated liquid handlers, or in-lab SDLs have comparable utility? This question is particularly dependent on the pricing model that these systems choose to employ.[12, 1] **Therefore, cloud labs will need to develop pricing models that are viable for academics as well as for industry.**

3.3 Reduced Flexibility

The incredible variability in materials, experimental setups, techniques, and methods has allowed researchers to push the envelope in an expansive number of fields but has also contributed to the difficulty in reproducibility seen in science today. Although cloud labs are poised to improve reproducibility, this is likely to come at the expense of flexibility. Robustness in the software and experimental practice requires tight control over how functions work, how they integrate, and how every detail of experimental procedures are performed. Experimentally, this robustness necessitates the broad standardization of procedures to ensure that experiments are performed correctly and without unexpected issues. Robustness also necessitates a cautious approach to incorporating changes to avoid disruptions or the introduction of unintended issues into the system. This is particularly true for cloud labs, given their complexity and the ambition to automate and integrate such a large array of procedures. As a result, researchers are no longer able to "duct tape" or "hack" solutions to problems like they are used to. They are limited to what is made available to them through the software stack. In this way, robustness is diametrically opposed to flexibility.
Reduced flexibility can also be faced by limitations in the software and hardware available. When specific hardware-software implementations support 90% of typically requested functions, researchers who use instrumentation in non-standard ways (e.g. materials scientists and nanotechnologists) can get left out. As a result, specific research may be heavily constrained or outright impossible to conduct. Without vast demand for such niche capabilities it is unclear whether large centralized cloud labs that target a broad set of applications would be willing or capable of accommodating the unique requirements of specialized research areas. In our case, necessary procedures like specific thermal ramps and custom buffers for agarose gels have posed unexpected yet significant obstacles. Some flexibility has been afforded in our experience, with the ECL being accommodating and implementing changes to functions that have introduced new capabilities needed for our own research. However, this may not necessarily be the standard. Moreover, it is dependent on the complexity of the request, with larger changes being significantly less likely to be implemented.
For research that does not necessitate specialized protocols or sample handling, the capabilities already present could be sufficient to transition to a remote cloud lab workflow. **Still, prospective users from fields outside of chemistry need to keep these limitations in mind and understand that an**

intrinsic lack of flexibility might prevent typical workarounds. This trade-off in flexibility is particularly important to note for those that are accustomed to building their own systems. When looking at protocols that one might want to perform in the cloud lab, it is important to carefully examine all aspects and check for compatibility with hardware and software. Otherwise, one could run into a similar predicament to our own, in which we assumed the ability to use different buffers only to realize that it was not a possibility in the cloud and there was no short-term way to hack a solution together.

In contrast to the previous challenges which can improve with further technology advancements, this decreased flexibility is an inherent property tied to the necessity for system robustness and to the ambitious targeting of a broad array of research areas rather than specific subfields. Flexibility has been a fundamental aspect of academic research that has allowed the exploration of ideas through the unconstrained adaptation and modification of methodologies and approaches. By restricting the scope of creative problem-solving, this limited flexibility could hinder the formation and exploration of new ideas. This could make large, centralized cloud labs incompatible with traditional academic research, especially in niche areas. Modifications to the underlying cloud lab approach or the introduction of new technologies that improve adaptability could mitigate these limitations and could enable cloud labs to support the dynamic and diverse needs of academic research. Otherwise, alternatives such as modular in-lab automation systems might be more suitable for the academic field.

4 This New Paradigm in Practice

Although significant challenges remain in this early stage of the technology, as commercial entities, cloud labs are incentivized to minimize these challenges to facilitate their adoption. As such, it is crucial for the academic community to recognize this potential conflict of interest and for early adopters to be prepared to face various obstacles. Moreover, if cloud labs aim to make inroads within academia, they must align themselves with its transparent and open-access ethos. This alignment involves openly sharing pricing models and publishing comprehensive benchmarking and validation data. Such data is crucial for accurately assessing the efficacy of these platforms and fostering trust among academic researchers.

Still, cloud labs are beginning to become a viable addition to scientific research. The number of available procedures and instruments could be sufficient for labs that primarily rely on commonplace sample handling and techniques to offload a notable portion of their laboratory work to the cloud, assuming that implementation issues can be avoided. Today, cloud labs such as the Emerald Cloud Lab and Strateos have reached a state where science can be accomplished and many of the benefits reaped.[21, 22, 9] Still, scientists looking to integrate a cloud lab into their research might want an insight into what cloud-based research looks like.

As one might expect, cloud labs dramatically shift everyday research practices towards a more computational one. The increasing prominence of coding in science has already driven a prioritization of the skill set.[23, 24] One might naturally expect cloud lab research to require a high level of programming skill. However, advanced tools that aid the writing of protocols reduce the barrier to entry far lower than one might expect. These tools make getting started more straightforward as protocols can be prepared without manually writing code. However, to advance as a user, a much deeper understanding of the cloud lab framework and the coding language is necessary. A programming background will help diminish this learning curve. Gaining these skills and knowledge will help accelerate writing protocols, analyzing data, and troubleshooting. This required level of expertise is expected to decrease with further development and as community-developed tools are shared.

Moreover, a perfectly robust cloud lab might obviate the need for users to be familiar with wet-lab practices. However, at this point the lack of system robustness all but requires the researcher to have previous understanding and experience working in wet labs. For one, although the software has many checks integrated that attempt to catch improperly written procedures and errors, it does not check nor attempt to teach users best practices. Similarly, prior understanding of likely points of failure in procedures greatly aids in troubleshooting issues that arise. This intuition and understanding is something that is much easier to learn when working in a traditional physical lab as these are the aspects that are inherently invisible to the users of the cloud lab. During the development phase of the technology, the need for prior expertise may reduce the applicability of cloud labs in academic laboratories, where students typically possess minimal lab experience and where it is still common for students, particularly in the life sciences, to lack coding knowledge. As the technology approaches maturity, it might sideline valuable learning experiences and could conflict with the conventional educational role of academic laboratories in workforce training. The effect of the technology will ultimately depend on the evolution of the skills demanded as automation increasingly permeates scientific research and on how scientific instruction evolves to accommodate this paradigm.

In this transition from a physical lab to an automated one, some might be worried that the development of cloud labs might make Ph.D. students and similar roles obsolete. Ph.D. students are the people who navigate the firsthand difficulties of research, find solutions to overcome the inevitable everyday problems that arise, and begin to fully drive the research as they develop as scientists. Although cloud labs have the potential to substantially bypass countless hurdles, research will remain fundamentally a problem-solving endeavor. In this aspect, the scientific practice will not change, but the focus will shift towards experimental planning, debugging, and data analysis. That said, it will be interesting to see how A.I.-driven research progresses and what effect that has on the community.

Lastly, many researchers rely on techniques that may prove challenging to integrate into automated systems. These techniques may be inherently difficult to parallelize, exhibit low-throughput or might rely on manipulations that are difficult to automate. **As researchers make the transition to automated cloud labs, it becomes imperative to explore alternative techniques capable of yielding comparable results.** While this transition presents its challenges, embracing this shift offers a high-impact opportunity for researchers to contribute to advancements in location-agnostic and high-throughput research methodologies.

5 Current Capabilities

The current experimental capabilities of the ECL facilities can be broadly characterized into sample preparation/purification and sample characterization. The more generally applicable the procedure or technique, the more likely it is to be available as an option. Towards the more basic end of the spectrum are equipment for liquid handling, which allows for manipulation of volumes as small as 2.5 nL via acoustic liquid handling, typical solid manipulation techniques, and all other fundamental wet-lab equipment generally applicable across applications. Purification and characterization techniques are more limited and may lack specific capabilities or alterations needed for niche applications. However, the numerous experiment types available still cover a comprehensive range of applications, particularly in protein, nucleic acid, and smart materials research. These capabilities include numerous types of electrophoresis, chromatography, spectroscopy, and light scattering assays. Notable available techniques have been organized in **Figure 5**. A notable gap in the capabilities are imaging techniques such as fluorescence microscopy or atomic force microscopy. This could be due to the lack of appropriately automated systems. In instances that require such techniques, materials produced in the cloud laboratory can be shipped to the user for further imaging. This workflow is essential for our application of DNA nanotechnology in advanced manufacturing and bioengineering.

Essential for reproducibility and transparency, data produced by experiments as well as comprehensive sample histories are automatically stored and managed by the software. This online storage is designed to ensure privacy through the implementation of Amazon AWS systems, where data is restricted to authorized users. However, time has shown that no system is infallible and can be susceptible to technology vulnerabilities or human error. As such, data privacy and security will remain key concerns for the cloud labs and will warrant continual scrutiny.

ECL has also recently announced that their proprietary Symbolic Lab Language will be open source for research use, facilitating the implementation and development of external software tools. This point will be particularly important for those looking to implement their own SDL or similar systems.

Beyond ECL, Strateos, formerly known as Transcriptic, has been the closest competitor in this space.

The company has similarly provided users access to remote automated laboratory facilities differing in a key aspect: whereas ECL has aimed to provide users access to a broad range of research instrumentation, Strateos' approach has focused on providing access to predefined workflows with a focus on small-molecule and biologics drug discovery, cell and gene therapies, and synthetic biology. However, the company has recently announced a shift in its focus toward the implementation of on-site cloud lab.[25] This realignment may signal a larger shift towards decentralized, adaptable, in-house cloud labs.
In general, it is clear that the current landscape of cloud lab technologies lacks healthy competition. This gap could not only stifle innovation in the sector, but also hinder the viability of the technology by perpetuating higher costs and limiting access. The introduction of more competition, and particularly open-source initiatives, is vital for the stimulation of the field and could expand the availability of the technology and provide alternatives that can better meet the tailored needs of a diverse range of fields.

5.0.1 DNA Nanotechnology

In our own work, we have aimed towards transitioning the manufacturing and characterization of DNA nanostructures to the cloud lab. This is a portion of the field that is particularly poised for automation as it is a frequent and time-consuming task. Particularly, the initial optimization to find the optimal conditions for the self-assembly of a novel DNA nanostructure is prime for automation. This automation would free up significant portions of time while enabling research labs to test novel DNA nanostructures more often. Moreover, it could increase accessibility of the technology and in turn lead to increased progress and interesting applications.
ECL currently has the capabilities that would be needed to relegate this task to a script. Sample models for DNA molecules are available which are used to define sample composition. Designed oligomers can be shipped to the facilities from the user or, in many cases, directly from the manufacturer. If needed, these oligos can be resuspended and consolidated into a mix. This mix can be combined with other necessary components, including buffers, many of which are already pre-stocked and ready to be used by the researcher. From there multiple approaches to characterization can be taken. Polyacrylamide gels can be used to characterize smaller assemblies (with custom loading buffers). Dynamic light scattering can be used to assess aggregation. Fluorescence, from intercalating dyes or specifically placed modifications such as fluorophore-quencher pairs, as well as light scattering and absorbance can be used to characterize formation kinetics.

6 Conclusions

Cloud labs hold incredible potential for accelerating research, increasing reproducibility, and increasing accessibility. With the benefits being so vast, it is not shocking to see increasing interest in the technology from both industry and academia. Moreover, the COVID-19 shutdown that grounded research across the world to a halt has exposed the fragility of our systems and the severe lack of robust backups, particularly in academic research. For researchers looking at the possibility of implementing cloud labs into their own research, it is important to understand that this technology is still in its nascent stage. With it comes many considerations that must be made to fully determine the potential efficacy of its implementation. Researchers must assess the level of parallelization they need and if they have requirements outside the scope of "normal" instrument use. Users must also keep in mind the difficulties that they are likely to face in this transition period as the technology develops. Still, it is important to note that cloud labs have reached the maturity needed for research to begin to be transitioned to the cloud. Increasing demand is sure to result in an acceleration in the development of the underlying technologies as well as the proliferation and implementation of such cloud labs across scientific research.
As a final note, we would like to emphasize that among its benefits, the unparalleled potential of cloud labs to expand open science may have the most pivotal and profound impact on the scientific community. The extent of this transformative influence, however, rests solely on how we wield the technology. **In order to support and grow a broad user base in cloud science, academic researchers need**

to make a genuine commitment to unfettered and transparent knowledge sharing. As a scientific community, the onus lies on us to seize this opportunity and harness the full potential of cloud labs. In doing so, we not only maximize our own benefit but also drive a monumental shift toward fostering open and collaborative research practices. To stand behind this, we have chosen to freely share the protocol code that we have developed to generate and characterize DNA Nanostructures at ECL, which will be uploaded to the CMU cloud lab GitHub community.

Acknowledgements
This work was supported in part by Air Force Office of Scientific Research award FA9550-18-1-0199, NSF CAREER award #1944130, Cystic Fibrosis Foundation award CFF003814XX222-TAYLOR, a seed grant from the Manufacturing Futures Institute at Carnegie Mellon University, and cloud lab thread support from the Mellon College of Sciences at Carnegie Mellon University. Finally, the authors want to thank both the Emerald Cloud Lab and the CMU Cloud Lab for their technical support as we transition towards cloud lab-based production and characterization of DNA nanotechnology.

References

[1] I. Holland, J. A. Davies, *Frontiers in Bioengineering and Biotechnology* **2020**, *8*.

[2] R. D. King, J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, A. Clare, *Science* **2009**, *324*, 5923 85.

[3] R. D. King, M. Liakata, C. Lu, S. G. Oliver, L. N. Soldatova, *Journal of The Royal Society Interface* **2011**, *8*, 63 1440.

[4] R. King, O. Peter, P. Courtney, Robot scientists: From Adam to Eve to Genesis, Technical report, OECD, Paris, **2023**.

[5] M. Abolhasani, E. Kumacheva, *Nature Synthesis* **2023**, *2*, 6 483.

[6] M. B. Holowko, E. K. Frow, J. C. Reid, M. Rourke, C. E. Vickers, *Synthetic Biology* **2021**, *6*, 1 ysaa026.

[7] N. Hillson, M. Caddick, Y. Cai, J. A. Carrasco, M. W. Chang, N. C. Curach, D. J. Bell, R. Le Feuvre, D. C. Friedman, X. Fu, N. D. Gold, M. J. Herrgård, M. B. Holowko, J. R. Johnson, R. A. Johnson, J. D. Keasling, R. I. Kitney, A. Kondo, C. Liu, V. J. J. Martin, F. Menolascina, C. Ogino, N. J. Patron, M. Pavan, C. L. Poh, I. S. Pretorius, S. J. Rosser, N. S. Scrutton, M. Storch, H. Tekotte, E. Travnik, C. E. Vickers, W. S. Yew, Y. Yuan, H. Zhao, P. S. Freemont, *Nature Communications* **2019**, *10*, 1 2040.

[8] M. M. Jessop-Fabre, N. Sonnenschein, *Frontiers in Bioengineering and Biotechnology* **2019**, *7*.

[9] D. A. Boiko, R. MacKnight, B. Kline, G. Gomes, *Nature* **2023**, *624*, 7992 570.

[10] B. W. Ong, J. B. Schroder, *Computing and Visualization in Science* **2020**, *23*, 1 11.

[11] M. Baumber, Benefits of Parallel Processing, https://www.matconibc.com/blog/benefits-of-parallel-processing.

[12] C. Armer, F. Letronne, E. DeBenedictis, *PLoS biology* **2023**, *21*, 1 e3001919.

[13] M. Baker, *Nature* **2016**, *533*, 7604 452.

[14] D. R. Boulbes, T. Costello, K. Baggerly, F. Fan, R. Wang, R. Bhattacharya, X. Ye, L. M. Ellis, *Clinical Cancer Research* **2018**, *24*, 14 3447.

[15] A. Mobley, S. K. Linder, R. Braeuer, L. M. Ellis, L. Zwelling, *PLOS ONE* **2013**, *8*, 5 e63221.

[16] CMU Cloud Lab · GitHub, https://github.com/cloudlab-cmu.

[17] LUMICKS – Unlocking single-molecule and cell analysis, https://lumicks.com/.

[18] N. Slamnik-Kriještorac, R. Van den Langenbergh, T. Huybrechts, S. M. Gutierrez, M. C. Gil, J. M. Marquez-Barja, In *2021 IEEE Global Engineering Education Conference (EDUCON)*, ISSN 2165-9567, **2021** 637–644.

[19] L. Tobarra, S. Ros, R. Hernández, A. Marcos-Barreiro, A. Robles-Gómez, A. C. Caminero, R. Pastor, M. Castro, *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje* **2015**, *10*, 2 69.

[20] C. M. Ionescu, E. Fabregas, S. M. Cristescu, S. Dormido, R. De Keyser, *IEEE Transactions on Education* **2013**, *56*, 4 436.

[21] T. S. Frisby, Z. Gong, C. J. Langmead, *Bioinformatics* **2021**, *37*, Supplement_1 i451.

[22] R. P. Goldman, R. Moseley, N. Roehner, B. Cummins, J. D. Vrana, K. J. Clowers, D. Bryce, J. Beal, M. DeHaven, J. Nowak, T. Higa, V. Biggers, P. Lee, J. P. Hunt, L. Mosqueda, S. B. Haase, M. Weston, G. Zheng, A. Deckard, S. Gopaulakrishnan, J. F. Stubbs, N. I. Gaffney, M. W. Vaughn, N. Maheshri, E. Mikhalev, B. Bartley, R. Markeloff, T. Mitchell, T. Nguyen, D. Sumorok, N. Walczak, C. Myers, Z. Zundel, B. Hatch, J. Scholz, J. Colonna-Romano, *Synthetic Biology* **2022**, *7*, 1 ysac018.

[23] T. K. Attwood, S. Blackford, M. D. Brazas, A. Davies, M. V. Schneider, *Briefings in Bioinformatics* **2019**, *20*, 2 398.

[24] A. G. von Arnim, A. Missra, *CBE—Life Sciences Education* **2017**, *16*, 4 ar61.

[25] Strateos Announces Strategic Shift to Focus on Customer Demand for On-Site Cloud Labs, https://www.businesswire.com/news/home/20230418006219/en/Strateos-Announces-Strategic-Shift-to-Focus-on-Customer-Demand-for-On-Site-Cloud-Labs, **2023**.

[26] E. Sperr, PubMed by Year, http://esperr.github.io/pubmed-by-year/, **2016**.

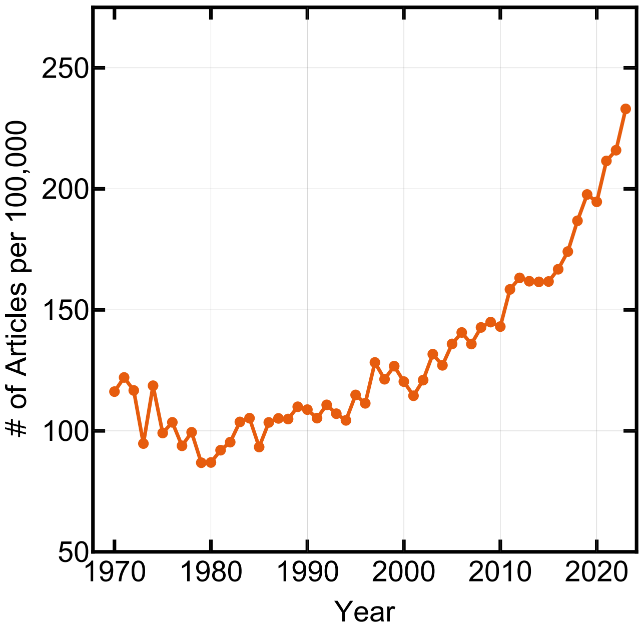


Figure 1: Number of articles per 100,000 indexed by PubMed containing the words "Automation" or "Automated" in their title from 1970 to 2023.[26] Note that the data for 2023 is incomplete as the year is not over as of the writing of this article. Adapted from.[1]

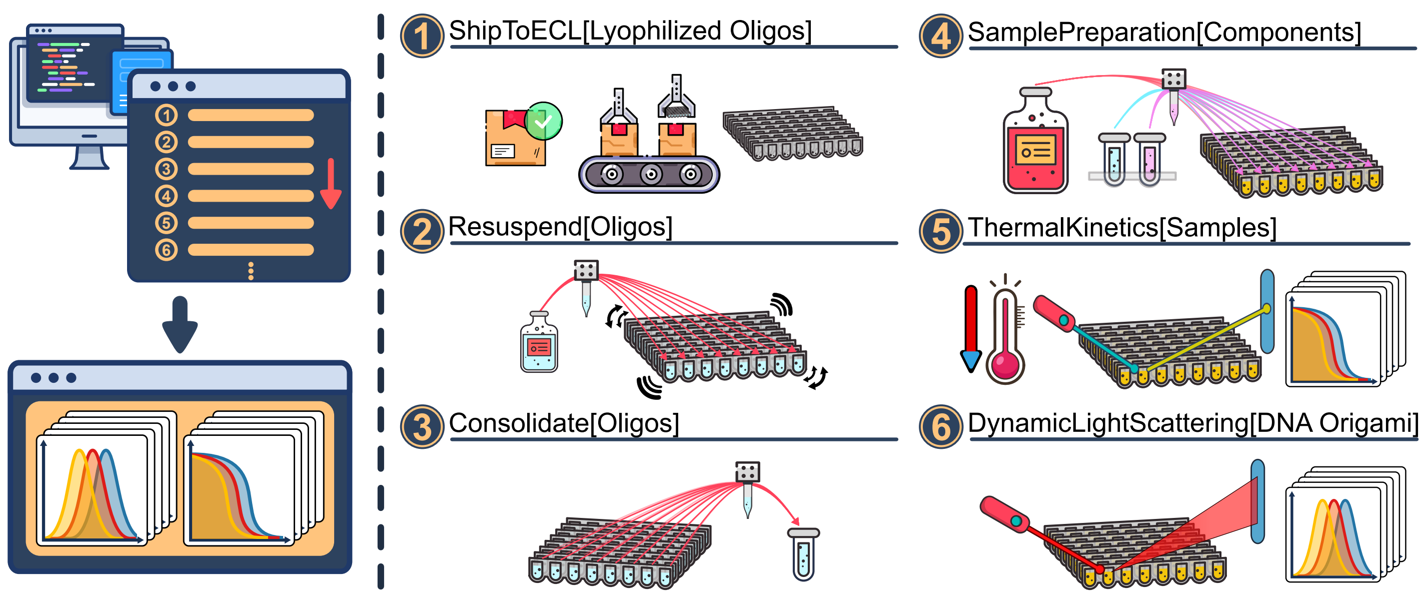


Figure 2: Illustration of an example workflow in a cloud lab laboratory. Left: The user writes programs that detail the desired operations. This program is verified and parsed by the cloud lab software and then uploaded to be completed in the remote laboratories. When completed, the data generated is uploaded for easy access by the user and for further analysis. Right: Each operation within a program is carried out in remote facilities using a mix of robotic automation and human operatives. In this example, necessary materials, such as the required lyophilized oligomers, can be shipped by the user or directly from a provider. These materials are then integrated into the system. The oligos are re-suspended in the desired conditions and consolidated to make the relevant mixture. A multitude of samples sweeping relevant conditions, such as ion concentration, can be easily prepared and then analyzed using relevant high-throughput methods.

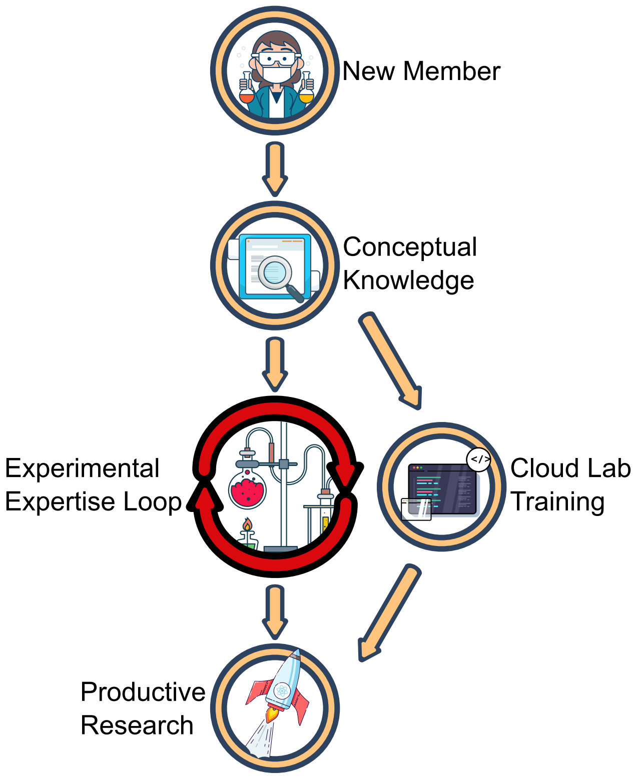


Figure 3: The cloud lab can help you bypass the time-consuming loop that new lab members have to go through to attain the necessary experimental expertise to become productive members of a lab.

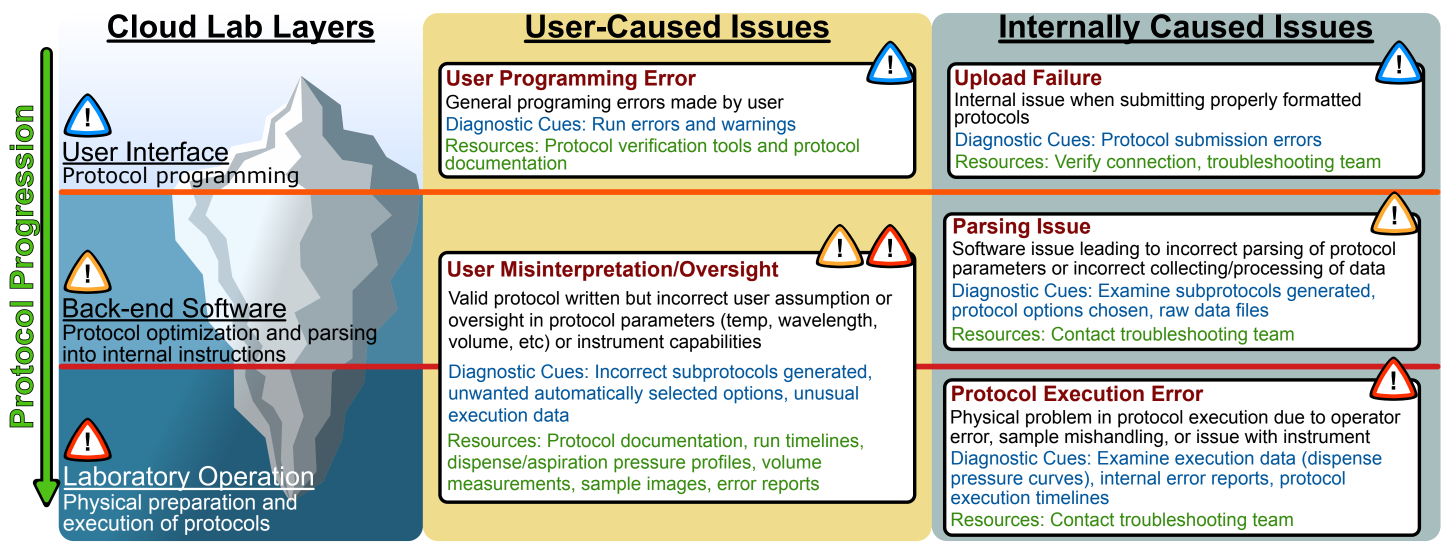


Figure 4: Left: The cloud lab can be seen as an iceberg where the user only has access to the top-most level: the user interface. This is what is used to interact with the cloud lab systems. In between the user interface and laboratory operations, where protocols are physically completed, is the software back-end. This intermediate software layer serves to manage, optimize, parse, and generally connect the user to laboratory operations. Multiple types of errors can occur as a protocol progresses from the user interface to the software back-end and finally to the laboratory operations. Middle: User-caused failures can occur at any step, with the majority due to simple programming errors that are caught by the software prior to protocol uploading. However, user misinterpretation or oversight can lead to protocols that are valid and proceed but do not generate the intended protocol or result in failure due to oversights such as not having enough sample volume. These issues are generally user-fixable and can be more easily diagnosed. Right: Failures caused by internal sources are more murky and often lead to issues that are hard to diagnose and generally require interaction with a troubleshooting team. They can occur at any step of the process, and do not generally raise an error flag. Implementations of fixes generally require an internal team as users do not have access to the systems that need to be modified.

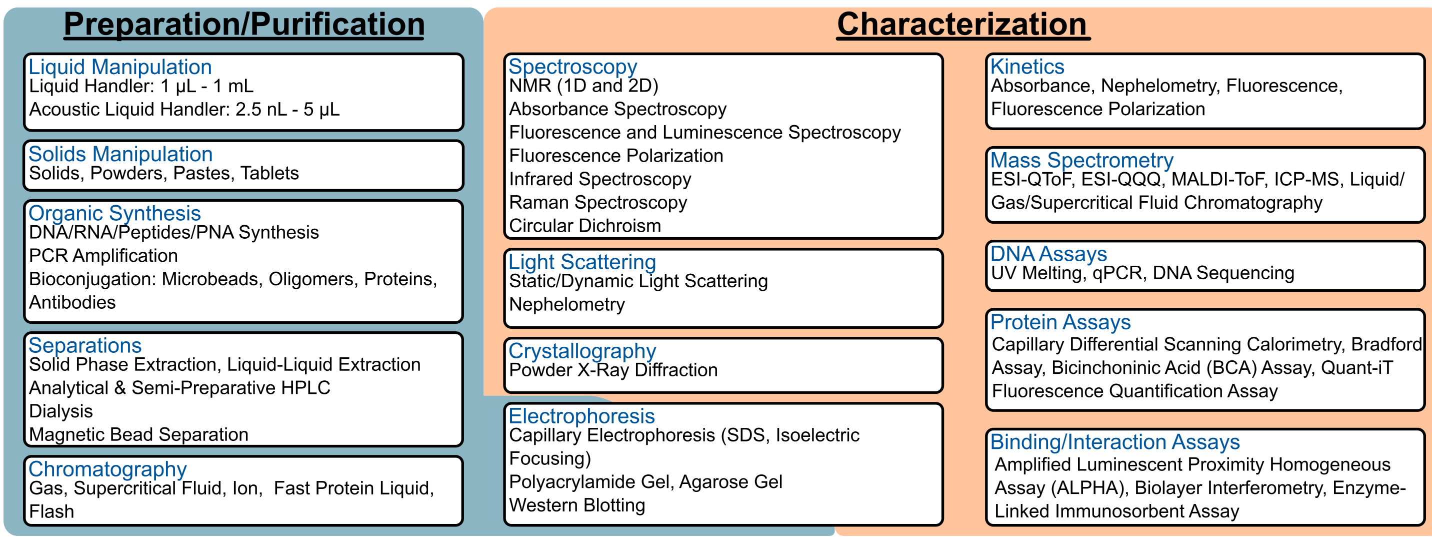
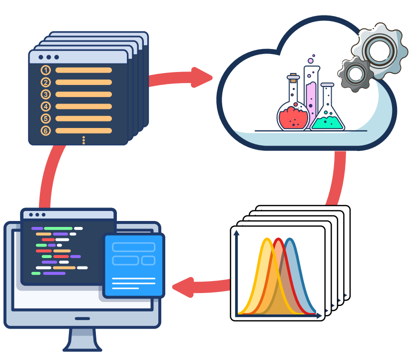


Figure 5: Overview of notable capabilities currently available at ECL facilities. They have been separated into sample preparation/purification and characterization sections, although many can serve multiple purposes. These experiments can be strung together in simple or complex programs as needed to run a series of procedures. More fundamental capabilities, such as sample weighing, are not shown but are available.

D. Sebastian Arias is a Ph.D. candidate in the Mechanical Engineering department at Carnegie Mellon University. He previously earned a Bachelors of Science in physics from the University of California, Santa Barbara. His graduate research is focused on the advancement of DNA nanotechnology with a specific focus on the development of novel DNA nanodevices and nanosensors in tandem with the development of cloud lab as a tool for nano-manufacturing.

Rebecca E. Taylor is the ANSYS Career Development Associate Professor of Mechanical Engineering, and, by courtesy, of Biomedical Engineering and Electrical and Computer Engineering at Carnegie Mellon University (CMU). Her degrees are in Mechanical Engineering with a B.S.E in 2001 from Princeton University and a Ph.D. in 2013 with Prof. Beth Pruitt at Stanford University. She joined the CMU faculty in 2016 and her lab now combines both microfabrication and nanofabrication to create hybrid top-down and bottom-up fabricated sensors and actuators for nanobiosensing, robotics, and advanced manufacturing applications.

Table of Contents



This forward-looking perspective highlights the transformative and burgeoning potential of cloud laboratory technology. It then assesses the technology's current and anticipated challenges, while scrutinizing its applicability to academic research. The article also provides essential guidance for researchers eager to delve into this innovative technology.