

## Part 2

# Surveillance of Wastewater and Prevention Platform (SWAPP):

# A Case Study of Digital Public Health and SARS-CoV-2

By Corina J Shtir, Austin D Swafford and Jeremy E Ellis

**This is Part 2 of a two-part series. For Part 1, see September 2021 issue.**

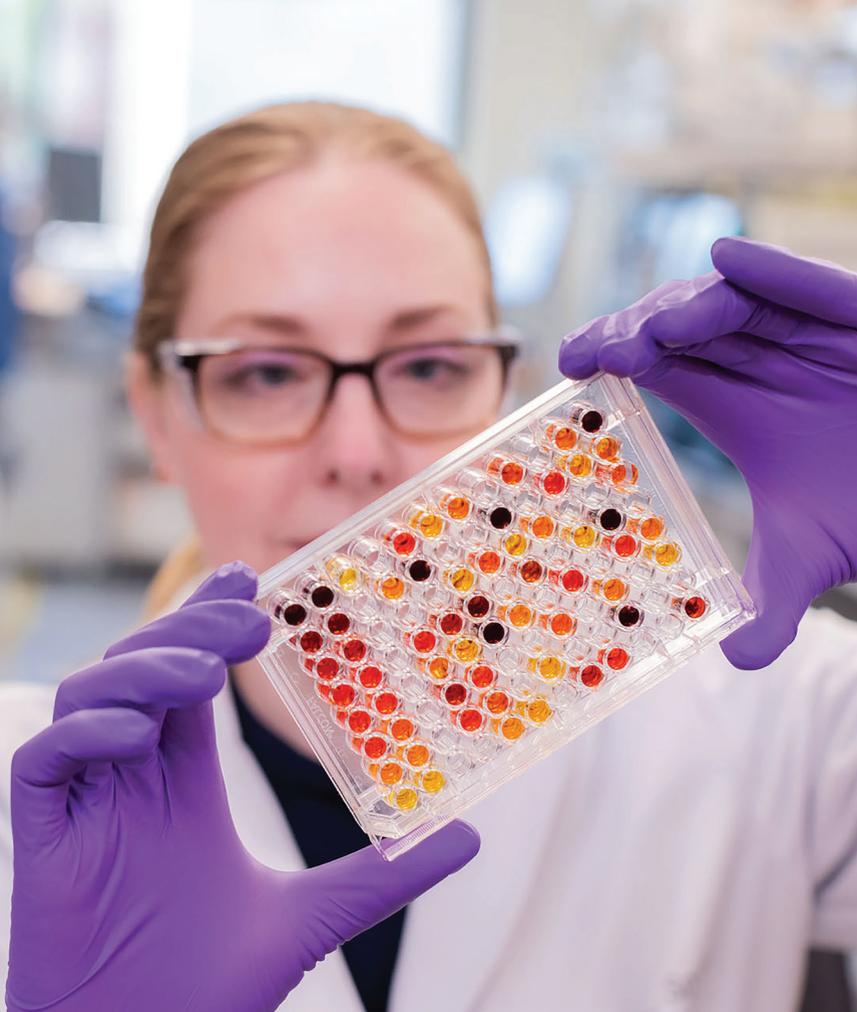
### Laboratory Perspective

The perspective of the testing laboratory is significantly more complex, as one or more tests are run and reported, with each test following different SOPs that may vary based on the report specifics provided by the wastewater company or by downstream customer requirements (Figure 3B). The combination of physiochemical

as well as biologic properties measured by laboratories by next generation sequencing (NGS), or high throughput screening assays can result in a high level of informatics diversity that must be accounted for when reporting results (Figure 4).

Unexpected variation in results may also occur if the number of samples surpasses the capacity of a specific laboratory; therefore,

testing must be distributed among different laboratories. The participating laboratories must follow standardized analytical protocol to avoid improper interpretation of results or variation in assay robustness. Unfortunately, laboratories do not frequently agree on a shared assay protocol, especially when results must be provided as quickly as possible and modifications to protocol



## Seeing cancer from every angle.

Our scientists are proud to lead the exploration of innovative, personalized cancer therapies to make long-term survival a possibility for more patients and the people they love.

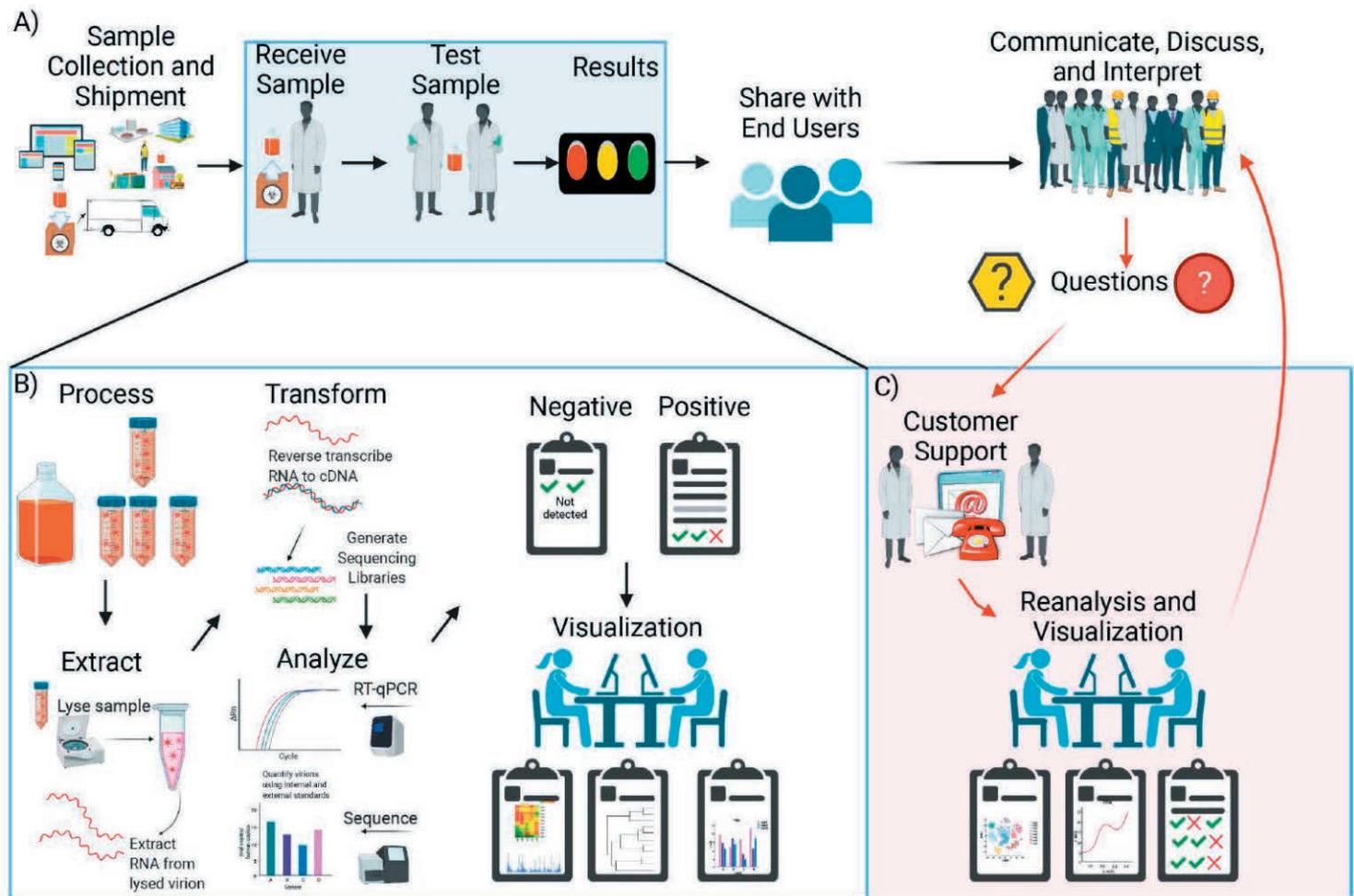
That's the touch that **transforms**.

 **Bristol Myers Squibb™**

Transforming patients' lives through science™

---

See more at **[bms.com](https://www.bms.com)**



**Figure 4: Costly Development Cycles for Laboratory and End Users**

A) A realistic overview of current wastewater monitoring efforts involves significant efforts of coordination and demands on the testing laboratory and end user organizations to properly visualize and interpret the results from wastewater testing. B) In addition to conducting the assessment of the wastewater sample, the laboratory will frequently be required to provide different sets of results and visualization based on positive vs negative results, evolving properties, and nomenclature for pathogens, and customer-specific contextualization. C) A non-virtuous development cycle may result where customer questions drive new analysis and visualization by the testing subject matter experts and organizations that subsequently impact the decisions and actions of the end-users resulting in more requests for analyses and visualization.

occur due to the urgency of reporting results. Moreover, labs differ in the degree of data and metadata capture, preservation, and monitoring, especially when dealing with data that is upstream or out of their work environment.

Automated QC monitoring provides a major opportunity for DH companies to support testing efforts, quality systems, and data reporting within and among laboratories. Furthermore, automated generation and updating of reports as a virus mutates and changes, together with automatic incorporation of additional demographic and environmental parameters and contextualization of findings, will cumulatively allow for standardized protocols. Standardization and quality control benefits from being able to make relative comparisons between results that originate from different labs, locations, or at different times. A digital health platform is essential for

this purpose, and this is where SWAPP primarily contributes by supporting automation and centralization of both data and results.

#### Customer Perspective

Customers must take countless factors into account when interpreting their data to make decisions that have profound effects on their business and the individuals involved in their operations. Some of the primary considerations may be detailed; however, these factors are highly dependent on the application and client specific details.

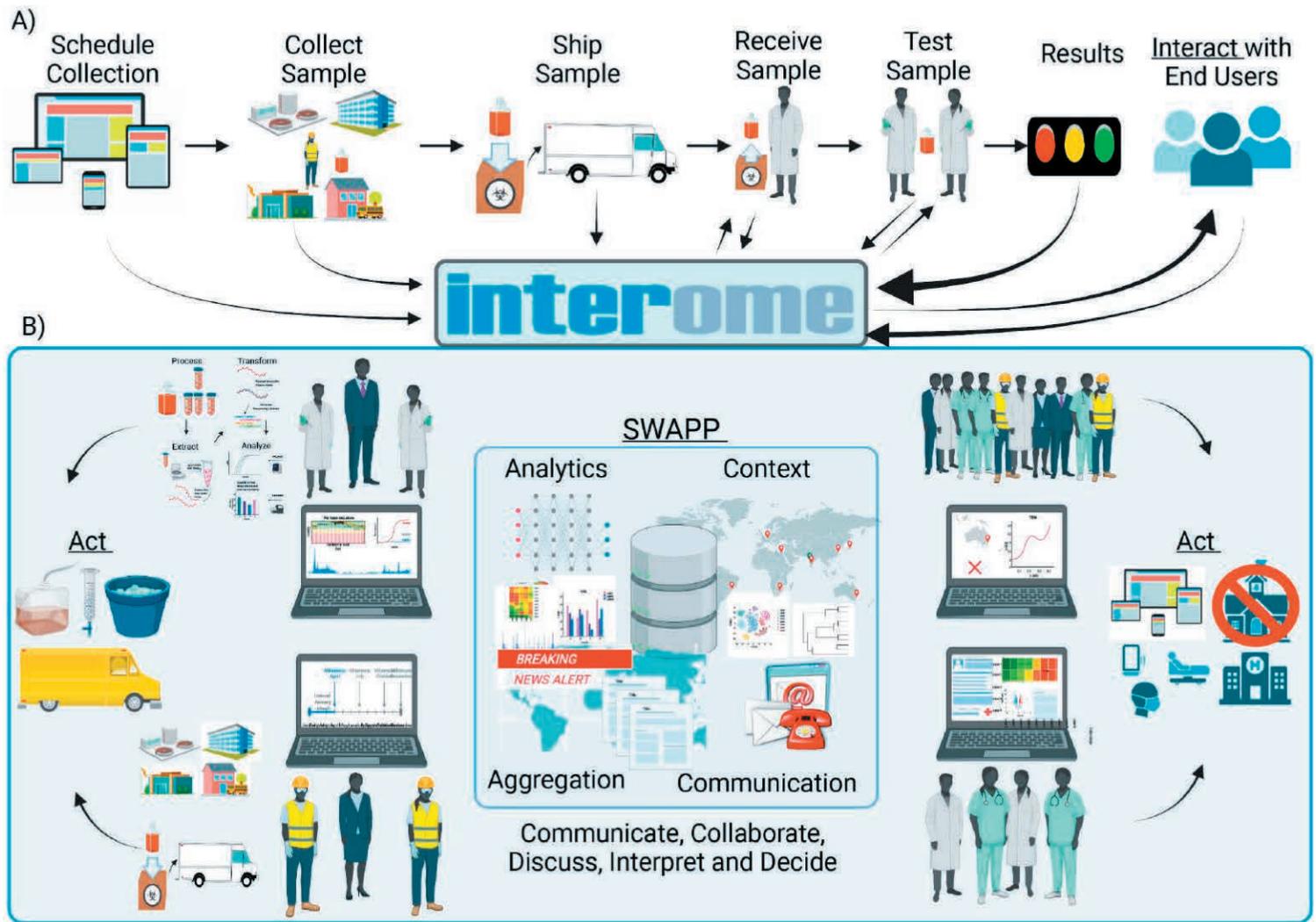
Data interpretation must be placed within the context of the result outcome and viral strain frequency and in context of the local region or wider geographic trends. If a wastewater test is positive, several additional challenges need to be addressed promptly and within the broader

context of their environment or demographics. Two primary considerations are:

1. how to accurately interpret the potential outcomes of the positive test, and
2. is the customer capable to properly navigate the result output format?

In addition, interpretation issues can occur when users are presented with a simplified report with a negative result and a complex report for a positive result. The way in which the information is provided and presented can create several challenges itself (Figure 3C).

Since PH efforts have not reached the full potential of their broad and standardized value (due, in part, to the scale and speed with which a pandemic occurs), end-users are often faced with having to develop their own



**Figure 5: Wastewater Testing SWAPP Solution**

A) A collection request to end-user workflow with InterOme's SWAPP platform at its core enables the idealized straightforward wastewater monitoring process to become a reality. Each component of the workflow can deposit and access the information needed for their portion of the process to inform decision-making for responding to adapting pathogen conditions and implementing quality improvements. B) Employing analytics across each component of the wastewater monitoring process while allowing for the contextualization of findings, the centralized aggregation of longitudinal spatially aware data and the ability to visualize and communicate findings enables SWAPP to serve as an effective digital health platform for enabling precision-public health quickly and easily.

visualization, description, or contextualization of data (Figure 3C). This burden is frequently shifted to the upstream data providers, i.e., the testing laboratories, which are then asked to engage in data analytics, visualization, and even interpretation that is not part of their standard operations (Figure 4B). Having provided newly formatted results to their customers, they then become subject to request to update, modify, re-visualize, and re-analyze results according to each customer's site and needs, all these tasks creating a feedback loop that can quickly become a drain on both laboratories and customer organizations (Figure 4C). Even if in-house trained engineers in either type of organization design a customized system to process and visualize results in a meaningful way, they are still required to develop ways to handle different

levels of information, changing variables, and to accurately and quickly determine if a viral source is moving from a sample with a single mutation to potentially several mutations that may need to be grouped into strains. End-users also face challenges when wastewater companies need to switch their preferred laboratory provider due to cost or strategic business reasons. New or different report types will require engineers to spend additional time engaging in data interpretation and science-related aspects rather than their initial or primary job function. In the effort to respond quickly to an adapting and evolving viral, political, and economic landscape, it becomes imperative to have access to local, regional, and global information that allows one to contextualize and update the resulting visualizations and models. This gap, combined

with lack of standards or adherence to standards in collecting, organizing, and reporting data and metadata are all risks posed to individuals, societies, and their shared health infrastructure, challenges which cannot be addressed by piecemeal efforts and non-virtuous development cycles (Figure 4).

In a highly connected digital world that relies on a global picture when comparing local data with a distant event it is imperative that automatic updating, integration, and interpretation become part of the WBE process, and that end-users have access to such resources. This burden is reflected in both individual organizations and governments; the impact of this requirement cannot be underestimated. All these issues require multiple functionalities via a DH platform, which we present in the following section. >

#### IV. SWAPP: A Digital Health Platform that Supports Precision Public Health Operations

A collaboration with Fry Laboratories

Big data analytics can have a profound impact on healthcare practice and systems by allowing all healthcare players, and most importantly health practitioners, to maximize the efficiency and utility of limited patient contact time. Electronic Medical Record (EMR) systems have changed the practice of medicine but are largely limited to centralized data aggregation and access. To date, healthcare systems are in the early days of integrating data analytics into the practice of medicine. In fact, the mass of unstructured data within EMRs is largely untapped and has the potential to be transformative using emerging machine learning (ML) and AI based technologies.

Data integration via analytics, artificial intelligence (AI), machine learning, and other computational systems are already impacting healthcare but must become a core component of healthcare that is intrinsic to routine assessment of wellness, prognosis, diagnosis, treatment, monitoring of disease, and preventive medicine measures. Intelligent data analytics and healthcare tools have already been adopted by radiology and pathology to improve diagnoses and prognoses, treatment response, and therapy choice. This is just the beginning of the broad adoption of such tools within medicine that will impact monitoring and management of wellness and disease. A healthcare system sustained by intelligent data integration across key healthcare players (hospitals, research institutes, biobanks, HMOs, individuals at large, etc.) will prove to be of even greater impact; it will save lives. Building an infrastructure that is flexible, scalable, and can sustain changes in data types or volumes of data, coupled with advanced algorithms that can analyze data within a regional or temporal context has resulted in the PH sector's growing reliance on DH solutions. WBE is but another example of the real-world value of DH solutions within the healthcare universe.

In the above sections we presented several core issues faced by multiple WBE service providers that aim to test the existence and spread of SARS-CoV-2 within a community. As described above, the stability of a virus is influenced by many variables. Without the ability to keep full-time infectious, genetic, and epidemiologic experts on call, given the global distribution of pandemic such as COVID-19, individual efforts are going to experience a rate limiting factor because of the inability to understand the severity and implication of the test results. Whether a result is inconclusive at a nearby site or whether a result is the first emergence of a highly pathogenic strain from a distant source, it needs

to be placed within the broader context of factors that contribute to the occurrence of an outbreak or emergence of a pandemic.

A large-scale digital health platform provides the medium for generating cohesiveness across all these data streams (**Figure 5A**); therefore, improving regional or large-scale surveillance and prevention of disease at the same time as safeguarding against system failures that stem from unknown or difficult to control variables. SWAPP has assembled a digital health infrastructure and system backed by constantly evolving advanced analytics, all of which can comprehensively support a diverse array of use-cases ranging from surveillance to prevention of disease for PH efforts. An additional advantage of being part of a DH platform such as SWAPP, is that the burden of managing data security, data storage, or data management is removed from the customers and managed by specialized systems.

The main elements of the platform include a centralized database, web applications, API services and data handlers, a sophisticated system for users and roles, dashboards that permit instant and live visualization of data results and QC metrics, alerts, and notifications for both data output and data quality, and a support hotline (**Figures 2 and 5**).

#### V. Brief Outline of SWAPP Components Three "I" pillars: Integrate, Interpret, and Intervene

The accumulation of multi-omic information negatively impacts the time doctors and healthcare practitioners need to internalize when deciding a diagnosis or treatment course. Surveillance and management of disease during COVID-19 pandemic offers a real-time example of how essential it is to link PM and PH into a unified approach that provides actionable information, which in turn can improve health assessment and management of disease. Our objective is to reduce complexity of fragmented data and create avenues that allow live integration of information, which ultimately should transform data into knowledge. Our three "I" pillars are built to *Integrate, Interpret, and Intervene* (**Figure 5**). Integrating difficult-to-access, unused, or unconnected information is one of the main gaps we deal with nowadays, without which digital health will not be able to provide profound assessment of wellness, disease, and treatment.

##### Centralized Database

InterOme ([www.interome.com](http://www.interome.com)) has established an infrastructure for data security, data management, and communication. In addition, our goal via our collaboration with labs and academic researchers is to develop and maintain a comprehensive

library of advanced analytics, backed by a central database and by data enrichment. This combination enhances the input data from labs and wastewater companies with public databases and other validated external datapoints. The combination of a rich centralized database supported by advanced analytics and API services facilitates integration, interpretation, and intervention across large populations but also for a targeted regional profile. For this reason, we setup a cloud database that facilitates the storage of data coming from test sites, and all other public or private domains that we can collaborate and work with. The centralized database tracks sites over time, is preloaded with GIS data, enables geographic data sharing, and aggregates statistics by census tract.

##### Main Backend

The SWAPP backend is designed to secure interactions with the centralized database. The backend system is built to host all APIs for the web and mobile applications, as well as all necessary APIs for labs and other platform collaborators. In addition, the backend system runs automated jobs, such as emails and text notifications, as well as automated data analyses.

##### Web Application

The SWAPP platform includes a secure web application that facilitates system, user, and data management and configuration. The web application provides access to data monitoring dashboards and reporting. To ensure high availability, the application has been architected with redundancy considerations.

##### Users and Roles

Access to data and dashboards is controlled by a privilege-based system that enables user configuration by site, by organization, or by region. A site manager can thus monitor the status of a site while an organization manager is able to monitor the entire organization and drill-down to the site level. Access to applications and data is secured by a dedicated Azure Active Directory B2C service. This provides user identity management and protects user specific data such as passwords, as well as the security authentication necessary for the integration through API services.

##### Application Programming Interfaces (APIs) Services

APIs automate the receiving of data from test labs or testing sites. SWAPP automatically converts reports into tabulated data while adding context. The APIs will be an 'always-on' connection to support all collaborators that either provide data or use data for analytics purposes. Also, a key

feature of SWAPP's design is flexible data handlers, which make the platform agnostic to the units or technology being provided, which, in the case of SARS-CoV-2 wastewater testing, is devised to support RT- qPCR, digital drop PCR, NGS, or any customized assay data. Through rule-based algorithms and natural language processing, data can be integrated in a standard reporting frame that will be easily connected with the SWAPP centralized database, which in turn will allow for automated updates of new gene references, mutation variants, or any external resources, all this offering contextualization of a finding.

### Dashboards

Real-time dashboards monitor and track the status of a particular metric or key relevant data points. Dashboards are also customized for particular use cases when needed. For example, a site manager could monitor the status of a particular site while someone at the organization level could monitor the entire organization, with the ability to drill-down to the site level. App-based dashboard interface enables visualization flexibility, clear positive/negative results, contextualized severity (type and quantity), intuitive user-customizable dashboard, and links to external sources (News, CDC/WHO, GISAIID, etc.). Moreover, dashboard interface enables visualization of QC results and contextualized changes (weather, seasons, holidays, etc.). Finally, dashboard interfaces allow companies to respond to dynamic customer needs without new infrastructure and analytics development.

### Alerts & Notifications

Users can setup their account to receive alerts and notifications by email and/or text messages. Alerts and notification triggers are setup based on key data points or metric thresholds. In addition, the system can notify sites when expected test data has not arrived.

## VI. Advanced Analytics

SWAPP is designed to implement automated advanced analytics to be run on demand, on-schedule, or based on the arrival of additional test result data. In this section, we highlight four approaches for advanced analytics that allow sequential building of complexity blocks, to ensure rapid operation, optimized scalability, and functional flexibility. We then follow with a brief description of the SWAPP main architectural components. As mentioned above, most processes necessitate the conjunction of multiple digital or analytical factors.

The main analytical sections that we work on continuously via our collaboration with labs and academia are precision PH via omics-based

approaches, explanatory and contextualized analytics, and prediction and forecasting models. Equally important, with aggregation of various datapoints throughout the journey of each tested sample, and from across multiple locations and during a period of necessary testing and validation phase, we will include automation for QC of metrics. This final step will contribute to the assessment of how SOPs should be adjusted to ensure efficient and uniform operations.

We note that across all types of services we work on, the ones that depend most on accumulation of time series, geospatial, environmental, or demographic data, are the explanatory and contextualized analytics, and the prediction and forecasting models. Also, as the healthcare sector is being integrated within the public, scientific, clinical, and individual space, we are already able to offer solutions that combine both PH and PM applications. To the point, the following section details the omics-based approach adopted within the SWAPP system for PPH analytics, with focus on COVID-19, via the detection of SARS-CoV-2, through our collaboration with the Fry Laboratories. Once those details are presented for context, additional layers of information regarding the SWAPP development process will be described.

Finally, we should emphasize that all these analytical categories are co-dependent, and that our continuous efforts to design sophisticated machine learning and advanced algorithms will leverage our centralized database. For brevity of describing prospective scenarios, we illustrate below some of the main benefits of the SWAPP design.

### Precision PH via an Omics-based Approach

Big data-analysis platforms, artificial intelligence frameworks, and any subsequent advanced analytics and decision management system exhibits the central reliance on both the quality and the quantity of the biological data and associated metadata. A PH example of such an approach would be the measurement of a single opioid metabolite concentration in wastewater samples via MALDI-TOF at various points across a metropolitan area. In this example, analysis methods would utilize the continuous analytical values obtained by the analyzer and combine this data with external geographical metadata to generate meaningful and predictive results. The resulting richness of the analytical results is largely constrained by the ability to collect metadata in addition to the relative scarcity of the primary data generated via the MALDI analyzer. As such, this type of problem naturally lends itself to an increasing multiplexity of analysis methods.

In this hypothetical opioid example, additional metabolomic compounds of opioids could be added to enrich the subsequent data used by the analysis method.

We contrast the above example with most medical decisions today, driven by a relatively small number of data points gathered sequentially over time from disparate sources (e.g., patient history, vital signs, lab values, imaging studies, physical examination, etc.). Recently omic-based approaches (genomic, transcriptomic, proteomic, and metabolomic methods) have become widespread, driven by the availability of massively parallel DNA sequencers and mass spectroscopy that can effectively and efficiently collect this data. While powerful, omic-based approaches have primarily remained as tools in the academic and research space, with some notable exceptions. NGS has been successfully deployed and adopted for cancer classification and treatment management (vs. mass spectroscopy widely used for detection of drugs and metabolites.) It is likely that these omic-based methods will continue to be more broadly adopted clinically, and their data integrated within the health care 'universe'. That in turn will require adoption of analytics capable of integrating that data into a comprehensive analysis of clinical value to health care practitioners.

Several factors contribute to the general barriers to adoption of omic-based assays. Simply put, most omic-based approaches utilize consumables that can be expensive per test. Justifying a higher cost of operation for an assay, when scaling existing single analyte assays often produces the same result, is a primary consideration. Specialized consumables, instruments, support equipment, personnel, and single-source reagents are all primary drivers of increased costs. Additionally, omic-based assays often are not available in a prepackaged form and must be developed in-part or in total by the laboratory performing the service. Lastly, omic-based assays often present an analysis problem as the volume of data and analytical meaning of the recovered results are not immediately tractable and require deployment of specialized analysis methods. To ease the transition between single-channel data collection methods and multi-omic-based approaches, the limited and efficient application of targeted omics may be useful. Too often the devices and applications that use omic-based methods are inappropriately separated into a false dichotomy with either single channel analysis methods separate from heavily parallel omics methods; however, this technology exists on a continuum of applications and approaches (**Figure 6**). Currently, most high-throughput service providers utilize multiplexed qPCR-based assays. ▶

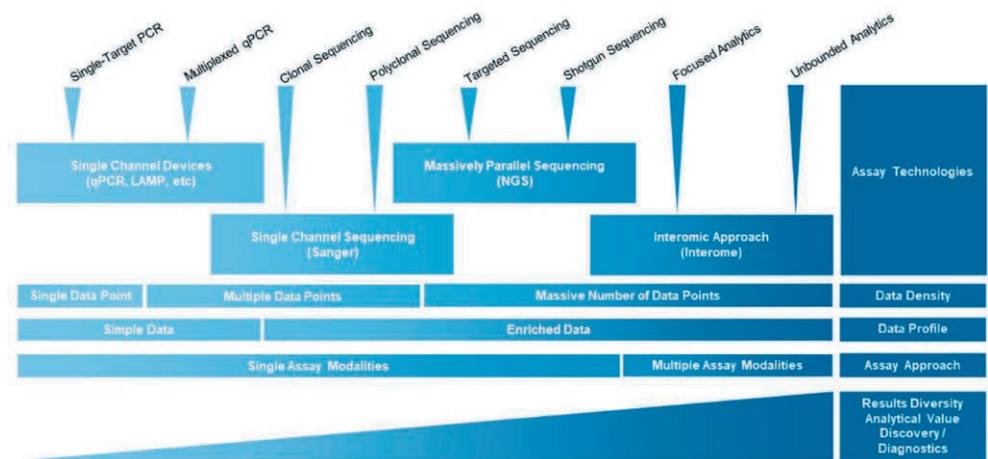
By selecting a device and application, it is possible to harness the massively parallel capabilities, extract additional data, but leverage efficiencies provided by the technology. For example, Fry Laboratories, LLC is a leader in clinically compatible targeted NGS-based sequencing for the efficient identification of prokaryotes. In this case, PCR is used to generate amplicons for specific regions found in the bacterial 16S rRNA loci by universal primers with the subsequent barcoded amplicons sequenced by NGS simultaneously with amplicons from other samples. This reduces the overall cost-per sample incurred using NGS, while still realizing much of the benefits of using NGS – the ability to capture the apparent diversity of the 16S loci in each sample. When coupled with specialized high-performance analysis pipeline the assay exhibits the benefits from NGS, but the robustness typical of targeted qPCR-based approaches. Furthermore, the use of such approaches may provide useful in bridging the cost-benefit gap often encountered by non-academic big-data analytic efforts.

### Specific High-value Application of Omics as a Proof of Principle

An approach to the successful deployment of advanced analytical platform could include a multi-phase approach, whereby the complexity of data acquisition, analysis, and flexibility are increased within each phase. It is proposed that a targeted NGS-based method, when coupled with a multi-omic approach, may provide a justifiable and targeted mechanism to support the use of an advanced analytical platform and provide demonstrable benefits. As described previously (Figure 6), NGS-based assays may be separated along a continuum from targeted sequencing to shot-gun or unbiased sequencing. In the simplest form, the gradient only describes the nature of the molecules that are interrogated by the device and do not describe other degrees of freedom for the assay.

An additional metric, in the case of NGS, is the analysis type as it may be an open or a closed analysis (Figure 7). A closed analysis will analyze the obtained data, either targeted or randomly sequenced, within a limited set of result possibilities. For example, an NGS assay analysis may only tally the presence or absence of a specific DNA molecule such as the detection of a prognostic BRCA1 mutation.

While useful, assays with a closed analysis may be used with advanced analytical methods, this is a quite different analysis type when compared to an NGS assay that performs taxonomic analysis of a wide diversity of molecules or that detects



**Figure 6: Nucleic acid-based assays and Result Diversity**

The existing and emerging technologies may be sorted by the assay result diversity with single target PCR-based assays representing a near-binary result as one of the most limited data sets. This trend may be extended to full unbounded integration of nucleic acid-based information with a plurality of omics-based assays.

novel mutations in various oncogenes. The density or richness of the recovered data per assay or encounter is significantly less than data obtained by open analysis methods. An example of an open analysis method used in conjunction with targeted sequencing is the 16S microbial identification assay, whereby the molecular targets of the assay are pre-defined portions of the bacterial 16S rRNA loci. The results may include an ever-increasing number of microbes deposited within the databases for identification. Furthermore, novel bacteria may also be detected, and the taxonomy may be derived based on sequence homology alone. In this way the possible set of results exceeds the initial assay conditions as the content (DNA sequence – base-per-base) rather than the presence or absence of a biomarker is the subject of the result. Given the state of the art of NGS and the current need to leverage omics meta-analysis, results and assay performance requirements may be more easily established by a targeted sequencing assay that generates informatically rich open results.

### Wastewater and SARS-CoV-2 Test Case

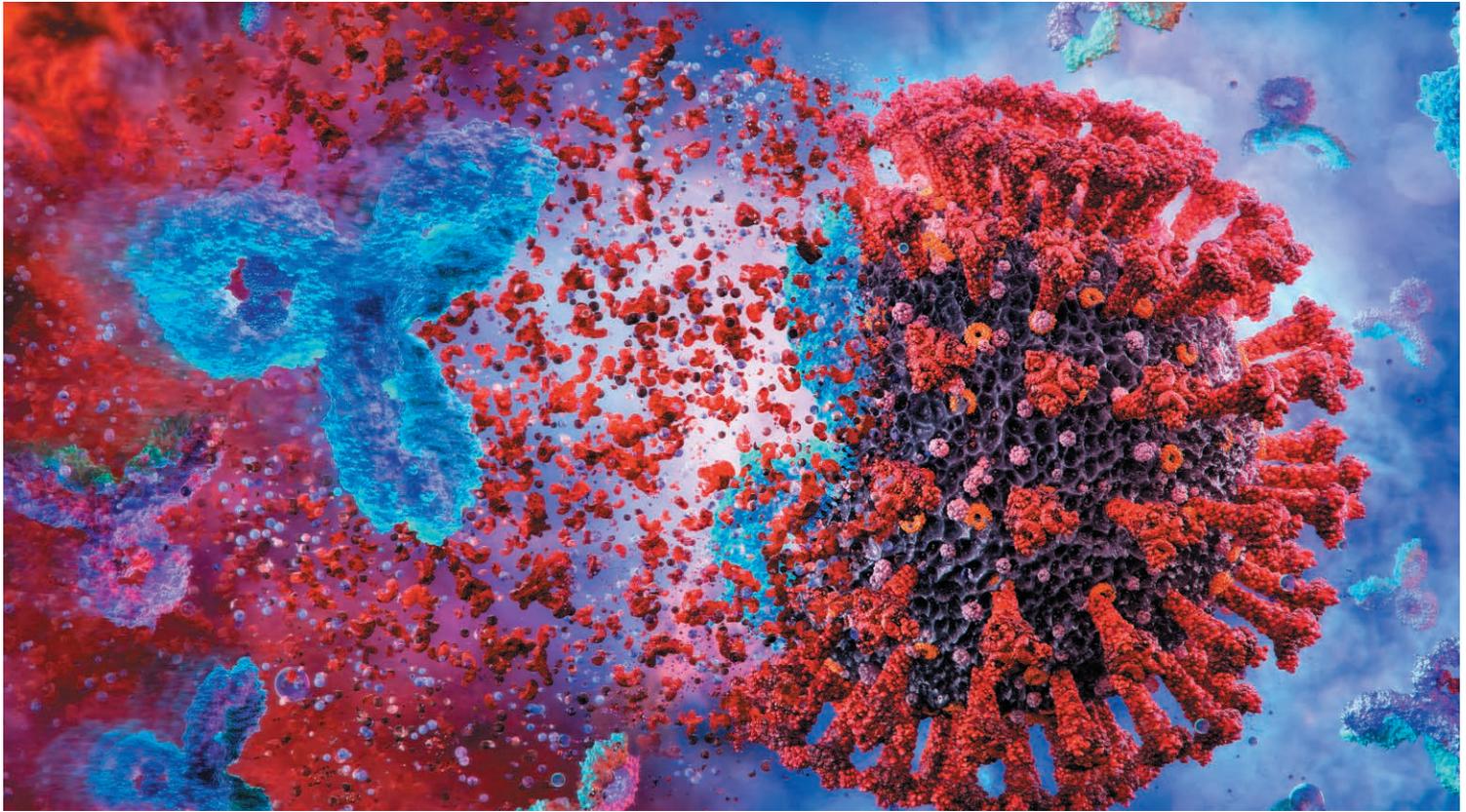
The current and ongoing SARS-CoV-2 pandemic exhibits a unique test case for the deployment of NGS-based assays and advanced analytics. Mutations in the spike protein of the SARS-CoV-2 virus are of relevance to epidemiology and PH. Fry Laboratories has developed several SARS-CoV-2 detection assays including an economical method to survey mutations of interest in the spike gene; in this method, the cost of the sequencing consumables is offset significantly with the number of samples processed. The assay allows for the simultaneous surveillance of

1,152 samples directly from complex, mixed, or contaminated samples.

While a useful tool that may be able to evaluate the relative risk of a population that contributes to a given wastewater sample for various spike protein variants, the benefit may have additional downstream consequences when linked to advanced analytics. Geographic data, population demography, and associated meta-data may be used in disease management and outbreak response. Increasing trend lines of a dangerous variant in a vulnerable population will allow for targeted application of intervention and protective measures to blunt the spread and impact of the virus. Additionally, advanced analytics may be used to combine known vaccination rates and variant information to target populations with specific vaccination types to ensure immunity. Lastly, data generated by this approach may be used to specifically route booster vaccines for specific variants to areas in need. It is expected that emergent and unexpected discoveries will become apparent that may assist in management of the virus.

### PPH Lab Decision Summary

Adoption of multi-omics or inter-omic-based approaches face similar or tougher challenges as deploying single omics assays. These challenges are cost of operation, return on investment, complexity of deployment, process management, and assay pipeline robustness. Targeted NGS-based assays offer a technological stopgap at a low cost per sample with robust operation. The approach yields significant quantitative and qualitative data that enable and feed advanced analytical pipelines. These features



allow for the construction of a meaningful data analysis toolset with demonstrable benefits that leverage the promise of both technologies. In contrast, attempting to deploy a fully open and agnostic analysis platform may be difficult to design and deploy without starting with at least some constraints.

We aim to realize precision health, and by extension precision PH, by seamless integration of big-data testing technologies and advanced inter-omic approaches. Precision health may in turn, improve the individual care of patients by harmonizing with individual screening, discovery, diagnosis, treatment, and outcome management. Additionally, in the case of precision PH, results may be realized in detection of disease causing or correlative agents and result in the real-time assessment of risk to vulnerable populations.

The SARS-CoV-2 pandemic represents a tangible, tractable, and motivating touch point where we aim to generate high resolution data that is amenable for processing and decision-making support for private, state, and federal epidemiologists. The next stage of the inter-omic approach is to deploy this assay in a collaboration between Fry Laboratories, LLC and InterOme Inc., to collect primary data and demonstrate the benefits of an advanced analytical and integrated PH option.

### Conclusions

The SARS-CoV-2 pandemic caused extensive disruptions around the world, affecting healthcare systems and global economies. A highly contagious and virulent virus combined with a wide range of reported symptoms across infected individuals continues to challenge the rate and utility with which incidence and prevalence data are collected and interpreted. The value of such data is further undermined by failure to detect and report seemingly unrelated symptoms and complications of infection, and most importantly, to identify 'hot spots' and 'super spreaders' that are driving the pandemic. Early detection of such events and locations could drastically curtail the pandemic. Measures taken to date have not been sufficient to do so, which has enabled persistence of the pandemic, with constantly emerging variants and surges globally.

The challenge is compounded by the inadequacy of individual testing to identify these events; in contrast, WBE analysis can identify both prevalence and emergence of new variants at a level of sensitivity and specificity that surpasses individual testing. This is illustrated by the fact that clinical test estimates are not representative of the full disease distribution across a community or at a given time, as many active cases remain undetected due to asymptomatic or mild symptoms; this failure in turn prompts lack of timely testing. Healthcare providers and public

health agencies are experiencing a burden that affects all aspects of life, including lockdown decisions and day-to-day personal or work-related actions.

Inability to test new cases efficiently at the community, national (e.g., CDC), or global level (e.g., WHO) continues to impact the assessment accuracy of disease trends. Inefficient testing, in turn, affects the measures that must be implemented in a timely manner throughout different regions or settings. PPH is a prerequisite for future management of disease and health, scalable for both small- and large-scale operations. Facing a formidable amount of data, workers in the field developed advanced analytics necessary for integrating complex and fast changing conditions.

The current world-wide surge in COVID-19 infections has catalyzed the evolution of digital health to sustain and drive precision public health efforts. Digital health enables PPH participants to collaborate and incorporate information across multiple locations and groups in real time. To the point, recent wastewater-based epidemiology provided real time, accurate, and cost-effective data that evaluates viral rates and infection trends across broad geographic areas and populations, while simultaneously mapping viral distribution spatially and longitudinally. In contrast, the clinical test framework has yet to achieve that level of performance. ▶

In this paper we summarized the operational elements necessary for supporting WBE screening of COVID-19, and illustrated the roadmap addressed via SWAPP, a dynamic, cloud-based digital health platform, developed by InterOme. We described the collaborative synergies between InterOme, private labs (Fry Laboratories) and academia, all of which are necessary for long term PPH management, at large scale and in real time. The workflow between wastewater companies, laboratories, and end-users forms a highly correlated matrix of variables – from water treatment, sample collection and shipment conditions to laboratory processing, and the inclusion of external datapoints. Each matrix variable influences the outcome and contextualization of information. The system defined by these processes characterizes the main role for which SWAPP is designed.

Recent WBE work emphasized the pressure healthcare systems experience and the need for innovation of advanced analytics applications for standardized SOPs, crucial actions for accurate mapping of viral hotspots, trends, and biomarker classification. These studies also highlighted the challenge we still face in adopting WBE methods, considering the small sample size of these cohorts and therefore a paucity of standardized protocols across global WBE efforts. While some wastewater indicators have been tested by a few research groups, many influencing variables remain unknown or need to be further investigated.

Understanding all aspects involved in the journey of a SARS COVID-19 sample may require wastewater companies to make changes in their protocols and laboratories to incorporate additional SOPs or address new demands. Nevertheless, WBE could provide a complementary solution for clinical testing by offering a cost-effective, time efficient, and scalable solution. As data across macro WBE ventures accrue, it becomes critical for SWAPP and similar DH platforms to expand collaborations with government and other private organizations to guarantee robust and rigorous SOPs are put in place. Also, future time series analyses will allow prospective and retrospective studies to evaluate which interventions maximized public health efforts, and under what type of conditions.

In addition to testing for viral load within wastewater or comparing WBE results with clinical reporting of individuals that may have been contributing to the outcome of a positive site test, we will further explore the use of concomitant testing for WBE antibiotic drug resistance. We note that while SWAPP is currently developed for the purpose of tracing and managing

		Assay Chemistry	
		Targeted / Amplicon Sequencing	Non-Targeted / Shotgun Sequencing
Informatics	Closed Results	Specific Target Molecules Restricted Result Output (Example: Cystic Fibrosis Assay)	Random Sequencing Restricted Result Output (Example: Chromosomal Trisomy Screening)
	Open Results	Specific Target Molecules Unbound Result Output (Example: Bacterial 16S Sequencing)	Random Sequencing Unbound Result Output (Example: Mitochondrial Mutation Screening)

**Figure 7: Assay Chemistry, Results Output, and Examples**

Next generation sequencing may be categorized along two differentiating features, being the assay chemistry approach and the overall informatics analysis pipeline. The combination of these two features is best suited to different application parameter space. Given the variability in these parameters a near term and attainable assay approach that is able to provide enriched data sets for integrated analysis systems can be identified by restricting the assay chemistry but leveraging the sophistication of open result informatics.

COVID-related activities, its infrastructure can naturally be expanded to test for other pathogens or toxic compounds. For example, the current pandemic exposed the vulnerability we face for a host of cases – future epidemics as well as the rise in meth and opioid usage, which further increases the urgency to identify opiate hot-spots and spreading trends, in young and vulnerable or low-income populations. Similarly, detection of lead poisoning and other toxins or viruses that are otherwise too expensive or too time consuming to address are natural areas SWAPP is devised to address. Lastly, we note that besides testing for human spread of disease, SWAPP is, by construction, relevant in other sectors, including aqua, plant, and animal farming sectors. In fact, the One Health approach specified by WHO recommends collecting information on viruses circulating in animals and the environment as a crucial step for the detection of future pandemics and for selection of potential human vaccines. SWAPP integrates current and incoming data by leveraging the intersection of biology, statistics, AI, and machine learning. The ability to examine the characteristics of a specific region that may not be represented by the average across its larger surrounding population, as well as the capability to construct the most accurate fitting explanation or necessary action to take, describes the future of DH. InterOme's long term vision is to generate precision public health knowledge through an intelligent infrastructure that creates a 'living' PPH

database based on which healthcare decisions can be optimized to support regional community, state, and Federal policies.

InterOme will continue to expand on SWAPP functionalities to generate advanced PPH prediction models for clinical applications that incorporate EMR information, longitudinal tracking of symptoms, and other phenotypic information. For example, SWAPP is currently focused on including genetic and phenotypic data to better stratify populations and to improve risk prediction for developing disease(s). As population specific COVID data accumulates, inclusion of applications built through hierarchical models that span diseases and their associated phenotypes will enable us to better understand various biological mechanisms, some of which are strongly correlated with COVID-19. For instance, long COVID-19 symptoms affecting cardiovascular disease (CVD) cases will allow us to reveal common pathways characterizing both long COVID-19 and CVDs. For this purpose, InterOme is collaborating with MyCardioGen, a new initiative in India, aimed at utilizing in the near term multidimensional diagnostic approaches, which in combination with InterOme's phenotype-based algorithms should improve population stratification of the Indian population, especially as it relates to CAD and associated phenotypes. Moreover, this effort will help validate clinically AI-based solutions by comparing their benefit versus standard-of-care treatments.

We intend to become a global healthcare tool provider that improves wellness and disease monitoring by efficiently integrating information that is otherwise unfeasible to track or may take years to process. The cost-effective aspect of DH-supported programs requires efficient prioritization of efforts and allocation of resources at the local and national level. Our goal via the SWAPP DH system and via our collaborations is to

set a groundwork based on which the healthcare community can optimize their decisions and practices.

Finally, we cannot overlook the fact that surveillance of population disease management without optimal communication mechanisms and support for improved behavior changes is not sufficient for successful implementation of prevention measures. DH systems will best be

fitted when behavioral changes at the individual and community level are incorporated within the routine management of PPH. The Wastewater industry can play a leading role within the PPH domain by becoming a foundation for healthcare players to establish and facilitate public health surveillance, improve risk prediction, and assist with generation of efficient disease prevention efforts. [GoDM](#)



### Corina J Shtir, Ph.D.

Dr. Corina J Shtir is CEO and Co-Founder of InterOme Inc, a global digital health company that aims to make complexity simple by developing and placing digital healthcare tools in the hands of healthcare practitioners,

for the purpose of enabling profound assessment of wellness, disease, and treatment. Dr. Shtir has a PhD in Biostatistics with applications in Statistical Genetics and Epidemiology, and (2021) Distinguished “Public Health Hero” Award by the US USC Keck School of Medicine University, (2013-2109, Thermo Fisher Scientific), VP, Global Head of Precision Medicine. Dr. Shtir led the global expansion of the company’s precision medicine platform and led the growth of Thermo Fisher’s Precision Medicine segment into an operation encompassing 200+ large or national-scale programs with a presence in 100+ countries.

#### Education:

2000-2002 MSc in Biostatistics

2004-2009 PhD in Biostatistics, Awarded Viterbi Fellowship

2009-2010 Post-Doctoral Fellow in Statistical Genetics, UCLA, US

2010-2012 Post-Doctoral Fellow in Statistical Genetics, Cambridge University, UK



### Austin D Swafford, Ph.D.

Austin D. Swafford, PhD, is Chief Technical Officer of InterOme. Previously, he served as director of Research-Innovations for the Center for Microbiome Innovation in the Jacobs School of Engineering at UC San

Diego. His expertise spans a wide range of fields critical to the advancement of science, technology and healthcare. He has led multiple private-public partnerships focused on automation, artificial intelligence, drug discovery, immunology, genetics, microbiology, and nanotechnology, as well as curriculum, business, and software development. His team supports numerous private-public partnerships examining the impact of commensal and pathogenic microbes, including SARS-CoV-2 on the health of humans, animals, and the environment as well as developing state-of-the-art methods, technologies, software, and bioinformatics approaches to characterize microbial communities. Dr. Swafford has co-authored peer-reviewed articles and patent applications on a wide range of subjects including basic and applied biology as well as novel sample processing and data analysis tools and methods to conduct large-scale microbiome studies. He is a member of the American Academy of Cancer Research, the American Academy of Allergy and Immunology, the American Society for Microbiology and numerous international collaborations including The Microsetta Initiative, the American Gut Project, the Earth Microbiome Project, the Global FoodOmics Project and The Environmental Determinants of Diabetes in the Young (TEDDY).



### Jeremy E Ellis, Ph.D.

Jeremy Ellis, Ph.D. is the Chief Scientific Officer at Fry Laboratories, LLC and BioD Genomics, Inc. Fry Laboratories is an independent clinical diagnostics and research laboratory based in Scottsdale,

Arizona; while BioD Genomics is a biotechnology spinoff company used to distribute innovative technologies developed in the clinical laboratory. Dr. Ellis is responsible for evaluating and translating high value biotechnologies to clinical use in addition to developing novel technologies that support next-generation DNA sequencing. He is an author of several recent patents regarding an ultra-rapid next-generation DNA sequencing technology and associated computational analysis methods. Dr. Ellis received his bachelor’s degree in microbiology from the Arizona State University and completed his Ph.D. studies in developmental and cellular biology at the University of California, Irvine. As an industrial scientist, he has focused on the development of clinically compatible comprehensive microbiology laboratory solution across all disease types enabled by cutting edge bioinformatics, targeted sequencing methods, and the sequence-based analytical definition of microbial taxonomy. More recently, Dr. Ellis has led the development and the EUA submission of a high-throughput SARS-CoV-2 detection assay using next-generation DNA sequencing methods. This assay development project is a result of an ongoing collaboration with Beckman Coulter Life Sciences.

### References (for both Parts 1 and 2)

1. Foundation, P.G.P., *WHY ARE AMERICANS PAYING MORE FOR HEALTHCARE?* 2020.
2. Chowkwanyun, M., R. Bayer, and S. Galea, “Precision” Public Health – Between Novelty and Hype. *N Engl J Med*, 2018. 379(15): p. 1398-1400.
3. Khoury, M.J., et al., *The intersection of genomics and big data with public health: Opportunities for precision public health.* *PLoS Med*, 2020. 17(10): p. e1003373.
4. Cevik, M., C.G.G. Bamford, and A. Ho, *COVID-19 pandemica focused review for clinicians.* *Clin Microbiol Infect*, 2020. 26(7): p. 842-847.
5. Wu, Y., et al., *Prolonged presence of SARS-CoV-2 viral RNA in faecal samples.* *Lancet Gastroenterol Hepatol*, 2020. 5(5): p.434-435.
6. Xiao, F., et al., *Evidence for Gastrointestinal Infection of SARSCoV-2.* *Gastroenterology*, 2020. 158(6): p. 1831-1833 e3.
7. Xu, Y., et al., *Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding.* *Nat Med*, 2020. 26(4): p. 502-505.
8. Barcelo, *An environmental and health perspective for COVID-19 outbreak: meteorology.* 2020.
9. Xagorarakis I., O.B.E., *Women in Water Quality. Women in Engineering and Science, in Women in Water Quality. Women in Engineering and Science.* 2020, Springer, Cham.
10. Hart, O.E. and R.U. Halden, *Computational analysis of SARSCoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges.* *Sci Total Environ*, 2020. 730: p. 138875.
11. Karthikeyan, S., et al., *High-Throughput Wastewater SARS-CoV-2 Detection Enables Forecasting of Community Infection Dynamics in San Diego County.* *mSystems*, 2021. 6(2).
12. Daughton, C., *The international imperative to rapidly and inexpensively monitor community-wide Covid-19 infection status and trends.* *Sci Total Environ*, 2020. 726: p. 138149.
13. Wurtzer, S., et al., *Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020.* *Euro Surveill*, 2020. 25(50).
14. Smruthi Karthikeyan, A.N., et al, *Rapid, Large-Scale Wastewater Surveillance and Automated Reporting System Enable Early Detection of Nearly 85% of COVID-19 Cases on a University Campus.* *mSystems*, 2021. 6(4).

### Acknowledgements

The authors would like to acknowledge the input, suggestions, and examples provided by Timothy Triche, MD, PhD, Professor of Pathology at the University of Southern California Keck School of Medicine.

### Part 1

Part 1 of this article was published in the September Issue of *Journal of Precision Medicine*.