EBRC Engineering Biology Research Consortium

Security Screening in Synthetic DNA Synthesis

Recommendations for updated Federal Guidance

A Policy Paper by the Engineering Biology Research Consortium Security Working Group[†]

Compiled and edited by Becky Mackelprang, EBRC Associate Director for Security Programs

April 2022

Introduction

The capability to quickly, accurately, and affordably synthesize DNA has greatly accelerated the pace of life sciences innovation and discovery. Research laboratories and companies can, with ease, acquire synthetic DNA products ranging from large pools of short DNA oligonucleotides to constructs that are several kilobases long. DNA synthesis and the techniques it enables are foundational to vast swaths of biological research and development, underlying advances in rapid vaccine development, accelerated progress toward renewable bioproduction, mitigation of the climate crisis, and the development of more nutritious and sustainable food. The same molecular research techniques, however, can also be used to cause harm. As such, research practitioners and policy experts have recognized a need for mechanisms that ensure responsible use. Many providers of synthetic DNA verify customers and screen ordered DNA sequences as important safeguarding measures, but these providers need updated guidance from the U.S. Government to ensure their screening approaches are as effective as possible.

Early DNA Synthesis Governance

While a segment of DNA from a pathogen is generally harmless by itself, it may be used by a skilled actor to reconstruct a virus or engineer additional or enhanced virulence factors into an existing microorganism. Considerable attention was turned to DNA synthesis screening in 2006 when a reporter from *The Guardian* was able to order a segment of smallpox DNA for delivery to a London apartment. Academic, industry, and government work and discussion in the latter half of that decade (e.g., Bügl et al 2007, Check 2006) resulted in two enduring steps forward. In 2009, five major providers of synthetic DNA[1] formed the International Gene Synthesis Consortium (IGSC) and developed a Harmonized Screening Protocol (updated in 2017) to support secure and responsible growth of the industry. The IGSC has since grown to include 23 members which IGSC. Providers of synthetic DNA must demonstrate screening capabilities to join the IGSC, although no retesting is subsequently required. Rather, members commit to adhere to the guidelines outlined in the Protocol for record keeping, order refusal & reporting, regulatory compliance, customer screening, and gene sequence screening using the IGSC Regulated Pathogen Database and internationally coordinated sequence reference databanks. In 2010, the US Department of Health and

[1] The five founding members of the International Gene Synthesis Consortium were Blue Heron Biotechnology, DNA2.0, GENEART, GenScript, and Integrated DNA Technologies.

Human Services (HHS) released the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA (the Guidance). The Guidance recommends baseline standards for screening sequences using a "best match" approach, where each 200 base pair span of an ordered sequence (and a sequence's six amino acid reading frames) are screened against large sequence databases. When the "best match" to the ordered sequence is unique to organisms or toxins on the Select Agents and Toxins list (SAT) under the Federal Select Agent Program (FSAP) or, for international orders, unique to an organism or toxin on the Commerce Control List (CCL), thereby falling under control of the Export Administration Regulations (EAR), the Guidance recommends what it refers to as "follow-up screening" before fulfilling the order. Follow-up screening, as defined in the Guidance, involves working with the customer to assess their authority to use the sequence and the legitimacy of their end use. The Guidance also recommends customer screening to verify customer identity and legitimacy and follow-up screening as necessary. While DNA (and RNA) synthesis capabilities and applications have grown considerably in the intervening time, U.S. Federal Guidance to providers remains the same as it was in 2010, leaving holes yet to be filled.

While the percentages of screened and unscreened commercially produced synthetic DNA are unknown, the IGSC suggests that its members represent approximately 80% of global DNA synthesis capacity. It therefore is likely that the majority of synthetic DNA produced globally is synthesized by providers that adhere to the Guidance and/or the IGSC Protocol, but a nontrivial amount of synthetic DNA is likely produced and shipped from companies that have no, or inadequate, screening mechanisms in place. Those companies that do not screen are at risk of inadvertently violating national regulations and/or international agreements by unknowingly shipping prohibited sequences domestically or internationally. They might also indirectly contribute to nefarious activities through the synthesis of non-regulated sequences with clear potential for misuse for questionable customers. For those that do screen, outdated Guidance leaves providers of synthetic DNA with a lack of clarity on key elements of screening (discussed in detail below). As such, individuals without the training, need, and/or authority to handle sequences that could be used to cause harm may be able to access and use them—intentionally or not—to damage human and/or environmental health and well-being.

Current Governance Challenges

Biosecurity governance systems are challenged by the international nature of research and commerce and inherent uncertainty around biological risk. Synthetic DNA is an essential research tool produced, shipped, and used all around the world. The drive to find new solutions to on-going health and environmental challenges fuels technological advances that also simplify the use of biotechnology to cause harm. Still, the successful deployment of a biological attack would not be straightforward. In addition to engineering or obtaining the biological agent, it would have to be grown to the appropriate scale, properly prepared for dissemination, and effectively dispersed. Acknowledging the tension between security and progress and building governance structures that balance these appropriately to achieve the greatest outcomes is a grand challenge.

If the U.S. had zero tolerance for biosecurity risk within its borders, it could impose strict restrictions on the production and distribution of a wide variety of synthetic nucleotide sequences. This limitation would come at the expense of maintaining leadership in biological and medical research. Policymakers and other stakeholders therefore have to grapple with how providers of synthetic DNA should screen customers and their orders, what they should screen for, how or if to incentivize and verify

screening, the types of sequences that should be screened for, the correct locus of control and/or responsibility within the government, and where liability ultimately lies for negative outcomes. In 2007, Garfinkel et al. described three policy intervention points for increasing security around synthetic DNA: commercial providers of synthetic DNA, owners of benchtop DNA synthesis equipment, and researchers who use synthetic DNA. 15 years later, some of the policy options suggested are still relevant. These include storing customer and order records, requiring a license to own a benchtop DNA synthesizer and to buy reagents for its use, and incorporating biological security curricula into researcher education. The structural challenges of 2007 also still exist. Who would be responsible for licensing purchases of DNA synthesis equipment? How and for whom is security education necessary? Furthermore, any policy decisions in the United States have to be made with an eye toward competition within an international DNA synthesis market and bioeconomy. Onerous restrictions domestically will delay U.S. research and make U.S. companies less competitive without containing international risk. However, given a lack of guidance in other countries, the U.S. has the opportunity to help set international standards and norms.

Governance is also challenged by the rapid development of DNA synthesis and assembly techniques. For example, the Guidance does not recommend the screening of oligonucleotide pools (pools of many, shorter DNA sequences generally less than 200 base pairs), which can now easily be used to construct much longer sequences. The Guidance also does not address bench-top DNA synthesis equipment: these devices pose a critical challenge to the sequence screening process described in the Guidance since they are meant to be operated behind customer firewalls. Myriad sequences that are excluded from regulation under FSAP and the EAR can be used alone, or in combination, to cause harm. Engineering biology is expanding the potential to design sequences that circumvent screening systems, for example by altering the genetic code that maps codons to amino acids. The Guidance specifically only applies to synthesis of double-stranded DNA, but single-stranded DNA and RNA can both readily be used to generate double-stranded DNA in a laboratory. Providers of synthetic DNA and manufacturers of gene synthesis equipment do not currently have guidance or regulation from the U.S. government for adapting their security measures to accommodate this expansion of technology.

Recent U.S. Federal and State Attention to the Production of Synthetic DNA

In 2020, HHS issued a Request for Information from the Office of the Assistant Secretary for Preparedness and Response on Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA, indicating that updated Guidance might be in development. Outreach from the U.S. Government to synthetic biologists and biosecurity experts to discuss this issue privately also demonstrates sustained government interest in more robust policy or guidance.

Given a lack of clear federal action and the pace at which engineering biology is growing, it is unsurprising that individual states have begun to invest attention in DNA synthesis security issues. Maryland House Bill 1256, which was introduced in 2021 and subsequently withdrawn, would have tasked Maryland's Department of Public Health to develop guidelines for gene sequence and customer screening by providers of synthetic DNA and manufacturers of gene synthesis equipment. Also in 2021, California Assembly Bill 70 (AB-70) was passed by both the Assembly and State Senate before being vetoed by Governor Gavin Newsom, who stated that "consideration should be given to whether a patchwork of state and federal regulations on biosecurity is the most effective way to approach an issue of international magnitude." AB-70 would have required gene synthesis providers and manufacturers of gene synthesis

equipment operating in California to conduct customer and sequence screening. Additionally, AB-70 would have required any entity receiving California state funds to purchase gene synthesis products or equipment from companies certified by the CA Department of Public Health (CDPH) as using adequate customer and sequence screening. New proposed legislation in California, AB-1963, is narrower than AB-70, requiring that California State University campuses only purchase synthetic DNA and DNA synthesis equipment from IGSC member companies, and requesting that the University of California campuses adopt the same standard (the legislature cannot legally require them to do so).

As AB-70 was under consideration by the CA State Legislature and Governor Newsom, there was an expectation among some that, given the size of California's economy and prominence in biotechnology, AB-70 would become a de facto standard for gene synthesis across the United States, similar to automobile fuel standards. Absent its enactment or public federal activity, there continues to be a need for timely, up-to-date DNA synthesis policy.

> *To bolster U.S. national security, we recommend the Federal Government: (1) update its Guidance for providers of synthetic DNA based on the advancement of technical tools and capabilities over the past twelve years; (2) develop and and support standards and metrics for evaluation of DNA synthesis screening systems; and (3) incentivize company screening practices by mandating that synthetic nucleotide products purchased with federal research and development dollars patronize companies with screening systems in place.*

Herein, we consider these three recommendations in depth, considering challenges to their implementation and opportunities to foster a safer, more secure DNA synthesis enterprise and research ecosystem.

1: Updating Federal Guidance on Sequence and Customer Screening

Two common approaches to sequence screening include block list approaches based on sequence of concern databases and best match approaches. In a block list approach, an ordered sequence is screened only against a list of known pathogenic, toxic, and/or highly concerning sequences. If the ordered sequence aligns closely with one of these listed sequences, the order is flagged for follow-up screening. This can return many false positives as sequences from benign organisms can be very closely related to sequences from the block list.

In a best match approach (recommended by the Guidance), a sequence is screened against a large reference database, such as the RefSeq non-redundant protein database. If a sequence is a best match to a sequence that is unique to a Select Agent or Toxin, the order is flagged for follow-up screening. Both approaches require designation of certain organisms or sequences as posing security concerns so that hits or best matches to those organisms or sequences are flagged. Currently, the Guidance focuses on the organisms and toxins listed in the SAT and on the CCL, but does note that i) pathogenic agents have many benign sequences (e.g., housekeeping genes) that do not play a role in pathogenicity and do not fall under regulatory control and ii) the SAT and CCL are not inclusive of all sequences with potential for misuse. Synthesis companies face significant uncertainty in deciding which types of sequences to screen for beyond these lists and need updated Guidance to address the screening challenges they face.

Before the end of 2022, HHS should develop and release updated Guidance for providers of synthetic DNA and manufacturers of DNA synthesis equipment that addresses the ambiguities of the 2010 Guidance, attends to technological advances, and describes a process for updating the Guidance every three to five years.

We describe four areas for particular attention in updated Guidance. *First,* the Guidance (appropriately) addresses screening for organisms and toxins on the SAT and CCL, but does not guide providers of synthetic DNA on the appropriate consideration of sequences with some homology to sequences from controlled organisms or toxins. It also lacks sufficiently detailed attention to sequences with security implications beyond those from listed agents. The level of harm that could be caused by such sequences exists across a spectrum. Determining which sequences are concerning enough that sequence screening systems should flag them is difficult and subjective. *Second,* at the customer level, it is common practice for research laboratories to have single accounts with many users, making true customer verification difficult. It is also common industry practice to sell through third-party distributors, complicating synthesis company intent to screen the final user. *Third,* flagged orders require time-consuming, and therefore costly, person-powered efforts by synthesis companies to verify the sequence or customer, so minimizing false positives is important. These efforts might ultimately be in vain if an order flagged by one company can be ordered from another company that does not screen sequences or customers. *Finally,* gene synthesis equipment (e.g., a benchtop synthesizer) is not legally required to have screening mechanisms built in and no synthesizers currently on the market have such capabilities. Some manufacturers are working on these capabilities, but neither federal Guidance nor the IGSC Protocol describe best practices or set minimal performance standards for screening sensitivity. Below, we describe these four considerations for improved synthetic DNA screening Guidance in greater detail.

Identification of Sequences of Concern

As researchers continue to expand our understanding not just of the role of single genes in model systems, but of complex genetic systems with context-dependent functions, and as the tools for organism engineering develop, it becomes increasingly difficult to scope the range of possible harms that given sequences or collections of sequences could be used to cause. Providers of synthetic DNA therefore face significant ambiguity in deciding which sequences beyond organisms and toxins on the SAT and CCL for which to screen. Some sequences would present a security concern only if introduced into a specific genomic and/or systems context. For example, human gut colonizing bacteria could nefariously be engineered to express human immune system modulators that overactivate or depress immune responses. However, the same immune modulating sequences needed for nefarious use are routinely ordered for legitimate research purposes. Screening for these types of security concerns requires attention to difficult cost-benefit analyses. It is impractical to screen for *all* potentially harmful sequences. A truly talented nefarious actor could find a way to misuse an overwhelming number of commonly used sequences. But the likelihood and consequence are small compared to the vast societal benefit resulting from technological advancement represented by DNA synthesis technologies. Other sequences—such as some of those from Select Agents—could be used without much imagination to cause much more catastrophic harm. The costs of screening for those sequences are relatively low when compared to the consequences of their misuse. However, such cost-benefit analyses are difficult because data is lacking and uncertainty is high. Determining a risk threshold above which sequences should be flagged, and then

identifying all the relevant sequences that fall above that threshold, will be subjective and experts will disagree. The level of concern of a sequence will also change as technology and biological knowledge develop, and thus cost-benefit analyses would need to be consistently revisited. The status quo, though, leaves ample room for the U.S. Government to contribute further guidance for strengthening screening systems without synthesis companies expending unnecessary time and money on flagging sequences with minimal potential for misuse.

In addition to determining if and how to screen for sequences that could be used to cause harm but are not from controlled agents or toxins, DNA synthesis screening also requires consideration of sequences that are unknown (do not exist in nature or have not been sequenced and deposited in a database), lack sufficient annotation, are from species with inadequate sequencing breadth to capture relevant sequence variations, and/or combine segments of sequence from different agents that, together, enhance pathogenicity. The prediction of function from sequence was not considered feasible at all in the past (Sequence-Based Classification of Select Agents: A Brighter Line) and today still presents a significant challenge. Tools that use homology to predict if a sequence might pose a threat, such as those emerging from IARPA's Functional Genomic and Computational Assessment of Threats (FunGCAT), could be incorporated into screening systems, particularly for use on un- or under-annotated sequences (e.g., Godbold et al. 2022).

Sequence screening must also be robust to sequence obfuscation. Sequences can be obfuscated by making extensive codon table adjustments and substitutions, shifting reading frames, ordering oligo pools that can then be assembled (the Guidance does not recommend screening sequences less than 200 base pairs), or hiding the parts of a concerning sequence within a larger sequence. Since the Guidance was issued in 2010, DNA assembly capabilities have grown significantly. Small (less than 200 base pair) fragments or oligos can be used to assemble much longer sequences. Some synthesis companies now offer RNA synthesis as well. Updated Guidance should incorporate these products into screening recommendations. Computational strategies can take advantage of next generation contig assembly algorithms to decrease the computational burden of oligo screening (Diggans & Leproust, 2019).

Customer Screening Guidance

Both the HHS Guidance and the IGSC Harmonized Screening Protocol call for screening of customers who order gene synthesis products. Like sequence screening, customer screening will never be perfect. But, the implementation of best practices can still prevent some negative outcomes, yield searchable records, and manage liabilities. Customer screening best practices include i) requiring customer identification data, including a physical, non-residential shipping address (PO boxes are not permitted given the lack of traceability); ii) restricted party screening to prevent any prohibited individuals or entities from receiving U.S. products or technology; iii) verifying a legitimate need—or, minimally, a scientifically sound use—and authorization before providing sequences for regulated pathogens or toxins; and iv) noticing "red flags" such as evasiveness around identity or affiliation. When any cause for potential concern is identified, companies should conduct follow-up screening, and if still unsure should have known contacts in the U.S. Government available to consult.

One challenge to customer screening is the growth of intermediate distributors ordering sequences on behalf of their end customers. Such distributors might, for commercial reasons, be hesitant

to provide their client (end user)'s identity, making customer screening by the synthesis company challenging. The U.S. government could explicitly recommend that synthesis companies require, in their distributor agreements, that distributors must provide end user identity upon request if ordered sequences raise biosecurity-related concerns.

A limitation to customer screening is the access of multiple individuals to single customer accounts. In laboratory settings, often a principal investigator or laboratory manager is the official customer, but may have many researchers ordering DNA through that account without oversight. Therefore, providers of synthetic DNA should not approve an order for sequence from a controlled agent simply because it comes from a trusted laboratory. To incorporate an additional layer of security, a synthesis company could require that an individual signed into a laboratory account enter their own unique passcode or PIN when placing an order. This would increase traceability within laboratories. Importantly, creating unique passcodes would need to be simple enough to minimize the urge to share passcodes with new or rotating laboratory members. Increasing the security of the ordering process can also make laboratories less vulnerable to hacking (Puzis et al., 2020). Alternatively, all orders could require approval from the laboratory principal investigator before being transmitted to synthesis companies. In reality, such a step is unlikely to provide meaningful security as the temptation for PIs to approve orders without reviewing them might be insurmountable. Another possible approach is for synthesis companies to send a weekly list to principal investigators of the organisms and genes with greatest homology to sequences ordered from their lab. This could be automated, would not delay the synthesis process, but would give PIs additional insight into lab activities.

The U.S. Government could also look to other industries with robust verification and screening methods. For example, credit card companies use aggressive algorithms to detect unusual spending patterns or purchases and flag them for fraud detection. A similar algorithm applied to DNA synthesis sequence screening could integrate the identification of anomalous orders with sequence screening. This could potentially reduce some false positives from sequence screening and increase sensitivity to orders placed over time by the same customer. However, such an approach would need to be taken with caution. In 2020, thousands of laboratories with no history of working on coronaviruses began ordering sequences for SARS-CoV-2. This could create a logistical nightmare for synthesis companies.

Challenges of Follow-up Screening

Whether a flag is raised by sequence or by customer screening, Guidance-adherent providers of synthetic DNA spend significant resources following up on flagged orders. According to a JCVI report from 2015, an estimated 5% of orders are flagged for review. "Yellow" flag hits might take one to two hours to resolve and "red" flag hits can take several hours. As the price of synthesis per nucleotide falls and the price of labor increases, the fraction of an order cost consumed by security screening increases, making it more difficult for companies that do screen to be cost competitive with those that do not (Diggans and Leproust, 2019; DiEuliis et al. 2017). Follow-up requires customer service personnel that are sufficiently trained to detect red flags during follow-up conversations, evaluate the validity of the customer's stated intended use, and make important determinations about order fulfillment and refusal. Such customer service personnel are generally more disposed to *help* customers and therefore may be hesitant to directly confront customers about their scientific goals and less experienced in detecting customer warning signs. Providers of synthetic DNA often operate internationally, so this customer service team must also be able to

communicate in relevant languages. Under current guidance, follow-up screening can involve attempts to verify a principal user's identity, making contact with relevant institutional biological safety officers or research directors, a literature review for related or relevant publications, and, if necessary, consultation with the Weapons of Mass Destruction (WMD) Coordinator at the nearest FBI Field Office. This highlights the importance of minimizing false positives in screening and of supporting tool development that enables small companies to implement screening systems without incurring an overly arduous financial burden (see below).

Currently, when follow-up screening reinforces the concerns that the initial screening raised, synthesis companies can decline to fulfill the order. That customer could attempt to order the same sequence from another company that might or might not screen. If screening were required of all providers of synthetic DNA, the customer would likely be flagged and denied again. However, because providers of synthetic DNA will continue to have some variation in their approaches to screening, and because a customer might more successfully obfuscate their order on subsequent purchase attempts, updated Guidance should consider if or how USG should encourage information sharing between companies. Currently, there is an information sharing mechanism in place within the IGSC, but is very rarely used. It is not known if its rare use is due to a concomitant rarity of concerning orders or if businesses are not comfortable sharing sensitive information with each other.

An alternative model that would take significant investment to establish but that might ultimately improve the safety and security of the synthetic DNA synthesis market is the establishment of an independent third party used by all providers of synthetic DNA operating in the United States for customer and sequence screening. Such an entity would need to be able to operate at the scale and on the timelines industry would require and could only be successful with industry buy-in and government backing. It would alleviate the problem of customers venue shopping for providers to ship sequences of potential concern and would make it more difficult to order the component parts of a dangerous sequence from different companies. The government could potentially share intelligence with the entity to inform their screening and surveillance efforts. The entity would need indemnification for its liability to not be untenably high. For the government to grant indemnification, standards and metrics to evaluate the quality and success of the entity would be necessary (see recommendation 2).

Implementing such a strategy would have some draw-backs, too. If a nefarious actor found a strategy to obfuscate sequences or otherwise elude the entity's screening approach, the actor could confidently order synthetic DNA from any provider of synthetic DNA. Furthermore, companies would likely be quite reluctant to send such extensive customer and order business information through a common entity.

Challenges for Manufacturers of Gene Synthesis Equipment

An emerging consideration not addressed by either the Guidance or the IGSC Protocol is the growth of benchtop DNA synthesizers. There are currently few (if any) synthesizers on the market that have built-in sequence screening capabilities, but there are some in development. Updated guidance should anticipate a future where these devices become increasingly common and identify and recommend appropriate customer and sequence screening standards.

If USG encourages or requires new equipment to have sequence screening capabilities, it should explicitly consider providing guidance as to i) processes by which flagged sequences can be approved for synthesis; ii) customer screening and export control during equipment resale on a secondary market e.g., through a unique synthesizer identification number; iii) safeguards against machine tampering (e.g., through verification procedures during reagent changes, inoperability in the absence of screening); iv) sequence tracking and logging, such as a "black box" of all synthesized sequences accessible to relevant authorities; and v) if screening is best done locally or through a "phone home" approach. Phone home approaches could create cyber vulnerabilities for companies or government agencies engaged in proprietary and/or sensitive research. On the other hand, a phone-home approach could help prevent local tampering. Reagent locking mechanisms and phone home serial number checks on consumables may help prevent some of these issues.

California's (vetoed) AB-70 did not require screening of gene synthesis products manufactured by an entity for that entity's own use. Thus, if a manufacturer of gene synthesis equipment sold equipment to a company, no screening would have been required for sequences synthesized for use within the company. Such a standard would leave a considerable vulnerability to insider threat. A large entity or institution, for example a university or a large biotechnology or pharmaceutical company, could have hundreds or thousands of unique users sending sequences to be synthesized, none of which would have been required to be screened under AB-70. To avoid building screening loop holes into updated guidance, the federal government should encourage screening of internally ordered sequences as well. If it does not, the government should advise manufacturers of gene synthesis equipment i) if machines purchased solely for use within an entity should/must still have the *capability* to screen (either locally or through a phone home approach) so that if machine ownership changes, screening practices can be maintained; ii) if customer screening is required for the sale of used gene synthesis equipment; and iii) if/how to verify when gene synthesis equipment is used to make products for internal use or for use by external researchers.

Manufacturers of gene synthesis equipment are constantly innovating and improving upon their products. Some are incorporating screening capabilities into future models. Prompt guidance from the federal government will help these manufacturers develop the most effective screening mechanisms, guide their approaches, and improve the uniformity of screening performance across the industry. Left without guidance, these companies could spend significant time and resources developing less-effective mechanisms.

2: Standards and Metrics for Assessing Screening Algorithm Efficacy

Many providers of synthetic DNA do screen the orders they receive for sequences of concern. Those that do generally make their algorithms robust to frame shifts and codon substitutions. Many screen for sequences beyond those from the SAT and CCL and work to make their algorithms resilient to more advanced efforts at sequence obfuscation. However, these companies lack systems to uniformly test screening efficacy. Developing such systems is challenging because they require standards to measure performance against. No such standards currently exist and creating such standards can be difficult, as demonstrated by these examples for which the outcome of screening is not obvious:

- A customer orders a sequence with "best match" to a gene from a Select Agent involved in pathogenicity, but with only 60% of amino acids in common.

- A customer orders a 1.2kb fragment. 300 base pairs of that fragment have 85% homology to an 800 base pair controlled open reading frame while the other 900 base pairs have high homology to a non-pathogenic organism.
- A customer orders a full-length controlled open reading frame but significantly alters a key functional domain and asserts that cited literature suggests the changes eliminate pathogenic function.
- A customer orders a plasmid for the expression of a gene from a human pathogen that exports virulence factors, but intends to express it in nonpathogenic bacteria.

Clear standards and best practices arising from Recommendation 1 would enable the development of metrics by which to measure screening performance.

> *To enable the developers of DNA screening algorithms to assess the efficacy of their sequence screening protocols, the Federal government should develop or guide the development of agile, quantitative screening performance metrics and methods and make them available for industry use.*

Metrics should incorporate rates of false positive and false negative findings against test sequence sets and evaluate capabilities for detecting obfuscated sequences. Metrics and evaluation standards will need to be regularly updated in tandem with regular (every three to five years) updates to screening Guidance and in response to evolving conditions.

The development of standards and metrics will require significant stakeholder engagement and technical expertise; if a technically straightforward mechanism for evaluating screening systems could easily be developed, it would already exist. At minimum, IGSC members, the National Institute of Science and Technology, the national security community, and security-minded engineering biology practitioners should be heavily involved. Developing and stress testing screening protocols and any developed metrics might necessitate some degree of biological red-teaming as nefarious actors may use sequence obfuscation or other methods to out-maneuver sequence screening methods. Concerns about information hazards can emerge from red teaming, but this practice also creates the circumstances under which vulnerabilities can be identified and attended to by good actors before they can be taken advantage of by nefarious actors.

3: Incentivization of Adherence to the Guidance

Currently, no incentives exist for customer or sequence screening in the United States. In reality, the expense of building customer, sequence, and follow-up screening capabilities creates an economic disincentive to screen. Moving directly from a Guidance model, however, to regulatory enforcement of screening might not give regulators sufficient time to develop rules with enforceable precision and clarity.

> *We therefore recommend that, after screening Guidance is updated, and no later than the beginning of FY24, the Federal Government incentivizes sequence and customer screening by requiring that synthetic nucleotide products purchased with federal research and development dollars patronize providers of synthetic DNA that screen sequences and customers*.

A major challenge to the implementation of such a recommendation is the establishment of a process to determine or certify that a synthetic DNA provider does screen and is eligible to receive orders funded by federal dollars. This would require a significant, ongoing commitment of time and resources by USG. Updated Guidance as described in Recommendation 1 and screening standards and metrics described in Recommendation 2 must be detailed enough to allow for the development of a certification process that enables Recommendation 3. Logistically, the requirement might need to be enforced at the funding-agency level. The bounds of screening requirements would need to be articulated with precision. Where high-precision parameters cannot be articulated, standards for certification should allow for dialogue between providers of synthetic DNA and a certifying body that demonstrates provider awareness and efforts toward robust screening. Because there are innumerable potential misuses of synthetic DNA, certification should not demand perfect screening around edge cases. As a starting point, we suggest that a certification program 1) ensure companies screen for sequences from organisms and toxins on the SAT and CCL and be robust to frame shifts and codon substitutions; 2) screen customers against the U.S. consolidated screening list; 3) develop mechanisms for screening oligos and pools of oligos; and 4) verify that providers of synthetic DNA maintain records for eight years.

A certifying entity would not necessarily need to be part of USG (e.g., could be the same entity as described in "*Challenges of follow-up screening*" above). A trusted third party intermediary could benefit from working hand-in-hand with both industry and government, while maintaining a degree of independence. Ideally, a certifying agency or organization would be able to support and inform updates to the Guidance in response to emerging conditions and quickly communicate new concerns to providers of synthetic DNA. If established with sufficient security mechanisms in place, such an entity could also become a home for communications from providers of synthetic DNA around problematic orders. The entity could potentially identify patterns across the synthesis ecosystem, and let synthesis providers know about concerns without compromising privileged information.

Alternatively, as suggested elsewhere, USG could require that synthetic nucleotide sequences purchased with federal funds patronize IGSC companies (see DiEuliss, Carter, & Gronvall, 2017). This approach would take advantage of the already established and widely adopted standards set by the IGSC Harmonized Screening Protocol and would obviate the need for a government certification process. However, there remain screening considerations for which USG guidance could be very useful. IGSC does not test members' screening capabilities after initial membership, so strengthening IGSC membership requirements, minimally through testing screening systems over time, would greatly increase the potential impact of such an approach.

Whether a DNA synthesis provider-certification program is directed by the IGSC, the US Government, or a third-party entity, it must be appropriately funded. As the function of a certification system would be to protect global health and well-being, it must be supported by public dollars with reasonable fees for providers of synthetic DNA. Providers of synthetic DNA that do currently screen do so entirely at their own expense, and high certification fees would add an additional burden for their responsible activity. Steep certification fees could also make entry to the market less feasible for start-up companies; thus tying fees to company size might be appropriate.

There are efforts to develop screening systems so that screening can be more easily implemented throughout industry. For example, the Nuclear Threat Initiative and World Economic Forum are bringing

together diverse stakeholders to develop an international Common Mechanism for DNA synthesis screening. An open-source tool called SeqScreen has recently been developed and is available for use. Efforts are being undertaken to reduce the number of false positives in sequence screening, for example by the Joint Genome Institute of Lawrence Berkeley National Laboratory, whose BLiSS (Biosecurity List Sequence Screening) pipeline can detect when flagged sequences are unlikely to be involved in the pathogenicity of an organism. Companies are also developing screening systems as a service. Battelle's ThreatSEQ screens for threat factors including antibiotic resistance, immune evasion factors, human bioregulators, protein toxins, and other threat factors (e.g., opioid enzyme pathways). This space is primed for expansion as companies built specifically around biosafety and biosecurity grow (e.g., Aclid). Current Guidance recommends that synthesis companies retain internal screening capacity and expertise. There could be situations, e.g., an early stage startup, where partnering with a trusted company, such as one that is an IGSC member and offers screening as a service, is a better option.

Important issues would need to be addressed before implementation of a certification system. For example, if a certifying or regulatory agency were given access to order records by providers of synthetic DNA to demonstrate their adherence to Guidance, it would be necessary to ensure that customer sequence orders could not become public through public records requests. Security measures should not compromise researchers' ability to conduct their research and protect their intellectual property.

Furthermore, any certification program should pay careful attention to legal liability. If standards and metrics are used for certification, those standards could be used to define company liability. This might incentivize providers of synthetic DNA to focus security measures only around those standards. If standards only focus on the most concerning sequences, others might slip through unnoticed. If standards are more broad, companies might respond by making sequences with very low levels of concern more difficult to access, thereby slowing research.

These and other challenges would need careful consideration before an incentivization and certification approach could be successful. Any consideration of transitioning voluntary, guidance-directed screening to a regulatory requirement would necessitate significant stakeholder input to ensure that the myriad implications were understood and addressed in advance. It would also be important to ensure that government expertise and resources were sufficient to keep pace with rapidly developing technology before establishing screening as a regulatory requirement.

Conclusions

Screening during the ordering process for synthetic DNA and DNA synthesis equipment is one important aspect of a safe and secure biological research and development ecosystem. Many companies pay laudable attention to security, investing significant resources into the development of screening and security systems and hiring individuals with the expertise to implement, improve and oversee these systems. As the capabilities of synthetic biology expand, it is becoming easier to predict function from sequence, assemble chromosomes and viral genomes, and use synthetic DNA to engineer organisms, consortia, and cell-free systems for specific purposes. These capabilities are cornerstones of our robust and growing bioeconomy. Concomitantly, these capabilities can simplify malicious activity by nefarious actors. The current approach to DNA synthesis security in the United States—suggested adherence to 12-year-old Guidance—is insufficient.

Herein, we have described important considerations for updating the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA, including improved sequence screening, customer screening practices, and follow-up screening on flagged orders and have discussed considerations for screening by manufacturers of gene synthesis equipment. We highlight the importance of screening standards and metrics and how they can be used to support a system that incentivizes providers of synthetic DNA to implement screening practices. These steps are among the most meaningful to support a safe and secure research enterprise and a strong bioeconomy that supports environmentally sustainable health and prosperity.

Funding Acknowledgement

This material is based upon work supported by the National Science Foundation under Award Nos. 1818248 and 2116166. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.