EBRC Engineering Biology Research Consortium

Public Comment

Draft Revised Guidance

Points of consideration from an EBRC Guidance Workshop

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

June 2022

Introduction

The issuance of proposed Revised Guidance for the DNA synthesis industry is laudable and of utmost importance to a safe, secure, and productive biological research ecosystem capable of addressing the great societal challenges of this time. EBRC hosted an NSF-supported workshop with key stakeholders across the industry as a forum for discussion and debate around elements of a robust Guidance framework. This "Part 1" submission responds to two elements of oligo screening that are central to the proposed Revised Guidance and that were considered at length during the EBRC workshop: 1) customer screening and responsibilities; and 2) sequence screening: sequences of concern. "Part 2" of EBRC's submission will address 3) sequence screening: technical considerations; 4) the sequence screening ecosystem; 5) benchtop synthesis equipment; and 6) emerging technologies. Risk cannot be eliminated from biological research and development, so we appreciate the care with which USG is working to balance the very real imperative to fervently support research with the implementation of security measures that are appropriate for the hazard space. We hope that our comment will be useful in the development of final Revised Guidance. Additional information and resources are available on EBRC's website.

1: Customer Screening & Responsibilities

The proposed Revised Guidance has an expanded scope that includes guidance for Customer, User, and Third-Party Vendor best practices, which would necessitate stakeholder education to enable and encourage compliance. Asking these stakeholders to invest their attention in security is certainly a worthy and positive endeavor that EBRC and workshop attendees broadly support. However, there was not unanimous agreement on how or by whom this education could be facilitated, the degree to which it might alleviate the burden on providers of synthetic DNA, and the security benefits.

Verifying customer identity and legitimacy

EBRC workshop participants highlighted areas where greater specificity in the Guidance could help providers determine customer legitimacy (which would also help customers submitting legitimacy documentation), as well as new approaches to determining customer legitimacy.

Clarification and specificity:

- The given definition for "*Verifying Legitimacy*" says that the recipient of materials should be a "legitimate member of the scientific community." Who does USG deem to be a "legitimate member

of the scientific community?" Examples where "legitimacy" might be unclear include: members of the DIY bio community; a customer who has not published research or presented at a conference in ten years, but now claims to work for a new biotechnology company somewhere in the world; a customer whose business address has the appearance of an apartment building.

- What does it mean for a Customer or Principal User to verify the legitimacy of an End User receiving a SOC? This has the potential to give license to someone in a position of power to ask a trainee personal questions. If an undergraduate has some past legal issues, are they disqualified from working with any SOC? Does a PI have a right to know that background and make decisions about a student's legitimacy? The Guidance could instead ask Principal Users or Customers to make sure that the End User has the requisite training and/or facilities to handle a SOC.
- Increase specificity and clarity on what documentation demonstrates legitimacy. Expanding on the examples already provided in the proposed Revised Guidance, a final version could include specificity such as:
    - A customer ordering a sequence that is on the CCL or Select Agent or Toxin list can demonstrate legitimacy through i) proof of registration or licensing with FSAP or DOC *AND* ii) documentation of an institutional biosafety officer.
    - A customer ordering an unregulated SOC can demonstrate legitimacy through i) documentation of an institutional biosecurity officer *AND* ii) providing relevant publication history *OR* supporting business license *OR* proof that a federal agency is funding the customer to work with SOC(s) in question *OR* sharing a description of intended use signed by an institutional biosecurity officer.

  Increased specificity would help DNA synthesis providers comply with Guidance, help customers submit appropriate documentation, support continuity across industry, and lend "top cover" to providers, whose customer interactions benefit when providers can support and explain their documentation requests with explicit passages from Guidance.

  A great challenge is providing the specificity and clarity needed to be useful to industry while maintaining flexibility for unique circumstances, such as customers changing their research focus in response to the emergence of a novel pathogen.

New approaches to customer legitimacy that could be enabled or endorsed by USG:

- Customer seal of legitimacy or "allow lists:" Some EBRC workshop participants favored increasing the relative focus on customer screening as compared to sequence screening, noting that sequences themselves are almost always benign; it's how a customer uses them that matters. Known or verified customers could be placed on allowlists or given a seal of legitimacy for given types of sequences for some amount of time, after which the customer would have to be re-verified. All sequences would still be screened, but follow-up would not be necessary for customers ordering SOCs for which they have approved uses. Implementation would be challenging, particularly on an international scale.
- Designating access tiers: Potential supporting documentation for legitimacy tiers are described above for customers ordering SOCs, while a customer who has been verified (has an address, has a payment mechanism) could order any sequence that is not a SOC.

Customer submission of legitimacy documentation during ordering

In suggesting that customers provide information to verify their own legitimacy when ordering SOCs, the proposed Revised Guidance has the potential to streamline customer screening and ease provider burden. This is a laudable goal, although would face some significant challenges:

- Companies would each need to build and integrate into their ordering platforms a mechanism for customers to easily submit this information as well as infrastructure to secure, track, and structure the information in a way that is useful to providers and eases customer and follow-up screening.
- Without substantial efforts to educate customers on why, how, and what documentation they should submit, companies might invest in such a mechanism only to have it go unused by customers. Putting the responsibility for education solely onto providers would nullify the intended provider benefits (efficiency, burden alleviation). Therefore, a concerted effort to educate customers may need to come from funders, institutions, or others. Otherwise, this practice is unlikely to achieve its goals
- Providers may find it challenging to interpret documentation submitted by customers.
- Providers have found that customers vary in how forthcoming they are with information to verify legitimacy, so providers might still need to do significant follow-up.
- Without a public, common definition of a SOC, customers will be unable to determine if they are ordering a SOC that requires verification of legitimacy. Customer education regarding SOCs (definitions, identifying, etc) would be needed.

Customer submission of legitimacy should only be incorporated into final Revised Guidance if industry supports it, if a plan for customer education is developed, and if greater specificity can be provided.

Third-Party Vendors

The inclusion of third-party vendors in the proposed Revised Guidance provides an important additional layer of security that alleviates a burden on providers of synthetic DNA. Previously, providers did not have a clear understanding of if/how to conduct customer screening or follow-up screening for SOCs ordered by third-party vendors. Encouraging third-party vendors to assume that responsibility was generally viewed by attendees as appropriate. While there may be a burden for some to develop necessary systems, it is important that they do.

Some additional clarity as to the definition of a third-party vendor would be useful. Workshop attendees were not in complete agreement about what “reformulation” of oligos encompasses, e.g. is all assembly and cloning included in “reformulation?” Specific vendor types that may or may not be third-party vendors depending on this definition include those that order a coding gene sequence and express it to make proteins, those that order oligos and assemble them into a gene, those that order DNA and package it into a virus, and those that order a gene sequences and clone it into a vector.

The definition of Third-Party Vendor might also be expanded slightly to include organizations that distribute synthetic oligonucleotide constructs. For example, plasmid repositories do not “order oligonucleotides from Providers,” but do have an important role to play in the security landscape and the Guidance would be strengthened by their inclusion.

It might also be useful to explicitly state that a Customer of a provider might be a third-party vendor, increasing clarity that a provider must notify a third-party vendor when screening identifies a SOC,

and third-party vendors would be responsible for notifying the originating customer. If "reformulation" includes the incorporation of synthetic DNA from providers into constructs, it might be appropriate for third-party vendors to conduct sequence screening themselves and incorporate the context of the entire construct, when applicable; however, capabilities for assessing genomic context are still developing and require dedicated attention and resources.

Tracking SOCs

The proposed Revised Guidance "recommends recording transfers of oligonucleotides containing SOCs from Principal Users and End Users to any other individuals not listed in the original order, such as through a Material Transfer Agreement (MTA) or another sample tracking process." While this is a laudable goal and there would be a benefit to tracking at this level, there would be challenges to implementation in academic settings. Academic institutions generally lack the tools, infrastructure, and staffing for universally maintaining consistent records for chemical and biological inventory management. Thus, in these resource-limited settings, tracking of SOCs might not become a priority beyond practices currently in place for MTAs.

It may be more feasible in academic settings for individual labs to track the SOCs they order. The consistent turnover and changing responsibilities within labs, especially those with small budgets that do not have an administrative support person, would challenge the consistent implementation of tracking SOCs, so education efforts would have to be consistent and robust to support such ongoing tracking.

Above all, the tracking of SOC transfer from Principal Users and End Users to other individuals would require Principal and End Users to understand which sequences constitute SOCs. A User might know their sequence is of concern if a DNA synthesis provider notifies them as such during the ordering process, as encouraged by the proposed Revised Guidance. However, without a consistently implemented definition of SOC, providers will likely have different interpretations and screening databases, and thus SOCs will be defined and tracked inconsistently. Additionally, the number of sequences currently housed in laboratories across the country (and world) that "contribute to toxicity or pathogenicity" is gargantuan, and they are inconsistently annotated. Developing tracking systems and incorporating such sequences into them is a very significant challenge and benefits may not outweigh the costs.

With these considerations in mind, recommending the tracking of SOCs might not represent the most effective approach to enhancing security concerns posed by SOCs, especially since regulated sequences are already subject to transfer regulations.

2: Sequence Screening: Sequences of Concern

Defining SOCs

Providing a description of what constitutes a SOC is perhaps the greatest challenge of developing meaningful Guidance. It is appropriate that USG encourages providers to screen for sequences beyond those from FSA and CCL lists, but the Guidance would greatly benefit from further clarity and specificity.

EBRC Workshop participants discussed the benefits of defining a robust and attainable "floor" for screening while providing guidance to institutions working to raise the "ceiling."

- A framework constituting the floor (minimal screening attainable or even expected by all) could be drawn from the 2010 Guidance: "The U.S. Government recommends that the sequence screening

method be able to identify sequences unique to Select Agents and Toxins; to meet their obligations under existing regulations, for international orders, screening should also be able to identify sequences unique to CCL-listed agents, toxins, and genetic elements." Setting an attainable floor is important because some effort to screen by a provider is better than no effort, and minimal efforts can be built up over time.

- A clear framework is needed to determine which sequences might be included under a ceiling (best, industry-leading screening approaches) regime. USG could develop such a framework for determining if a given sequence is a SOC and socialize that throughout the synthetic oligonucleotide industry, specifically, and the emerging biotechnology industry more generally. In doing so, USG might consider that:
    - A vast number of sequences can "contribute to toxicity or pathogenicity" threatening to "Public health, agriculture, plants, animals, or the environment." For example, all antibiotic resistance sequences, which are used in most aspects of molecular biology, would be of concern under this definition. So would Bt genes, which are expressed in millions of acres of corn and cotton and are highly toxic to specific insects. *Thus, to avoid the dilution of follow-up screening efforts by excessive flags for SOCs, it may be appropriate to narrow the given definition of SOC, or provide additional guidance for which SOCs necessitate follow-up screening.*
    - To help bound which sequences are of concern, one EBRC workshop participant described using functions to define SOCs, e.g., sequences that cause direct damage to a host, that subvert host immune responses, or that enable host invasion, dissemination or adherence (http://doi.org/10.1128/iai.00334-21). The same participant suggested that, for greater specificity, USG could 1) determine host taxa that are most valued/cared about/necessary to protect; 2) identify the pathogens most concerning for those taxa; 3) document the sequences that enable pathogenesis for the identified pathogens. This approach, or more broadly listing priority functions of concern, would enable those pushing the ceiling higher to focus their efforts on the highest impact sequences. Annotating and curating sequences based on those functions would be arduous, but previous efforts demonstrate that with investment, this can be done.
    - The Guidance should recognize that many sequences that are not directly pathogenic or toxic can be used to cause harm. For example, significant progress is being made in the expression of opioid biosynthesis pathways in yeast (see https://doi.org/10.1007/s11101-019-09644-w). While production levels are not significant enough to influence the illegal drug market, such capabilities may soon be within reach. Likewise, enzyme pathways that synthesize harmful non-peptide toxins are also not directly harmful, but could foreseeably be engineered to produce dangerous toxins.
    - List-based approaches to bounding SOCs are limited to what is currently known to cause harm. To future-proof sequence screening, USG should encourage the development of homology and/or functional prediction approaches that could be used to determine whether a previously unknown sequence constitutes an SOC (e.g., http://doi.org/10.1128/iai.00334-21).

- The final Revised Guidance should be intentional about the extent to which it focuses on sequences causing mass harm vs those that may cause harm to individuals. Both are important and have far-reaching implications on public perception and acceptance of biotechnologies, but differentiating might facilitate the determination of screening floors, ceilings, and/or other tiers or designations.

Database development

The proposed Revised Guidance suggests the development of a database of SOCs. Developing a widely adopted SOC database would require significant (and ongoing) time, effort, and resources. For widest adoption, such a database could be developed and maintained by an international organization independent of USG but be recognized by USG as being compliant with Guidance to enable US provider adoption. USG direction on the following considerations would be crucial to database development:

- Tiers: A low tier comprised of FloorSOCs would enable non-screening companies to quickly implement screening against sequences of greatest concern. This tier would contain SOCs for which the source toxins and agents are well known, enabling transparency that can maximize uptake and support peer review/testing. The transparency and availability of this database tier would support norm building across the industry. Other tier(s) might have more restricted access, containing an expanded repertoire of SOCs with great capacity to cause harm that have not been published or that might not be obvious to a non-expert. These tier(s) could also include synthetic sequences designed to help catch obfuscated sequences or functional patterns. The engaged community of providers using this tier could also work toward its constant improvement. Providers might find it useful if Guidance suggested appropriate follow-up for different tiers, i.e. there may be tiers for which no follow-up is necessary, but that might be useful to store as a yellow flag (e.g., transcription factors influencing SOC expression, toxin biosynthetic enzymes).
- Thresholds: Sequences exist at every step between the extremes of clearly dangerous and clearly benign, so USG guidance around appropriate database SOC thresholds would be very useful for database development.
- Agility: A database might be more widely implemented if users are able to adjust parameters. For example, a SOC from an agent that is endemic in a given region may require less scrutiny there than in regions where the agent is absent. Level of concern of a given sequence will also be application specific, e.g. gene drives vs viral assembly. A database must also be rapidly adaptable to emerging threats.
- Housekeeping genes: A workshop participant suggested that even housekeeping genes from regulated organisms might raise concern, not because of the sequences themselves, but because of the inferences that might be made about the customer's use. The participant suggested that there are few reasons to order housekeeping genes from pathogens unless one is trying to reconstruct a pathogen genome (see http://doi.org/10.1089/hs.2020.0004). False positives could be minimized if providers truly limited their screening to the single best match. Others suggested that housekeeping genes from pathogenic agents should not raise concern and rather should be kept on an "allowlist" to avoid triggering provider follow-up.
- Genomic context: Many sequences are concerning only in given contexts, e.g., when other proteins are present to form a complex or complete a pathway or network. Annotated SOCs within a

database could incorporate information on the context effects that increase concern. However, significant research is needed in this area (more in EBRC Response Part 2).

- Information hazards:
    - Some workshop participants expressed a belief that databases should be 100% open considering that:
        - sequence information sufficient to start a global pandemic is publicly available.
        - the threat space is *vast*; unsequenced SOCs abound in the global environmental ecosystem and
        - pathogenesis and toxicity are context dependent: with a given host, environment, and expression level, all kinds of sequences can become concerning, therefore trying to hide all concerning sequences does not reduce threat but does limit opportunities for testing and improving the database.
    - Others felt strongly that some or all database information should be restricted, describing how a fully automated, highly secure screening system that is sensitive to sequence function, that encrypts sequence and customer information, and that is freely available to companies could enable rigorous screening without generating information hazards.
        - Providers expressed that a database with such access restrictions would complicate decision-making and communication for flagged orders. Customers generally do not appreciate the follow-up screening process, and this interaction could be further strained if a provider is unable to say why an order was flagged.
        - A need was also expressed for biosecurity evaluators to see the sequence annotation and any contextualizing information from the database to accurately assess the concern of a flagged sequence. In addition to better customer communication and decision-making, this also supports database curation as false positives are identified and understood.
- Database security: Ideally, an international organization responsible for the database would be free to determine and implement appropriate, but not excessive or prohibitive, security measures. A database is not of use if providers cannot access it and if it cannot be kept up-to-date with input from many stakeholders.

It's important to note that many providers have made significant investments in their own screening systems and databases and would be unlikely to adopt something new. Actionable guidance should therefore be available to enable such providers to ensure the robustness of their own unique systems.

3: Sequence Screening: Technical Considerations

Technical details

- Expansion from dsDNA to include ssDNA and RNA: This is feasible, implementable, and welcome guidance that will increase security.
    - With respect to RNA, it might be useful to consider that not all synthetic RNA is used to synthesize proteins. Small interfering RNAs (siRNAs), CRISPR guide RNAs (gRNAs), micro RNAs (miRNAs), etc, are active molecules by themselves that can affect DNA sequence and/or gene expression within an organism. Segments of these RNAs that are designed to bind to complementary target sequences would point to intended use, but such segments

can be as short as 17 nucleotides and contain surrounding elements (e.g., PAM sequence) that would complicate screening. In considering how to approach screening for these RNAs, the difference in security concern between replicating sequences in an organism's genetic code and sequences with other activity might be considered.

- Smaller screening window for non-batch orders (from 200 to 50 bp): At the EBRC workshop, there was broad, general consensus around the *capability* to detect 50 bp SOCs; however there are significant tradeoffs in asking providers to do this. A 50 bp window will increase the required compute power and increase false positives, thereby increasing provider costs. State of the art screening platforms can minimize these burdens, but those come at their own costs. Overall, USG should balance the risk posed by 50 bp sequences with the capacity for provider follow-up and the tolerance of customers for follow-up screening. If the number of false positives increases, provider follow-up will decrease in quality and customer tolerance may be diminished, potentially pushing customers to non-screening providers. If the risk posed by 50mers is high enough, then perhaps a 50 bp window is appropriate. If 50mers do not pose enough risk to warrant the burden on customers and providers, this window should be reconsidered. Quantification of these risks could be undertaken and funded by USG to provide better data for making these decisions.

Batch orders

Workshop attendees indicated that the attention given to batch orders in the proposed Revised Guidance is appropriate to increase detection capabilities for small fragments that can then be assembled into longer SOCs.

- The specific encouragement to screen constituents of batch orders of oligos can be approached in at least two ways:
    - Combinatorial screening with state of the art technology: This approach yields significant and increasing numbers of false positives as screening windows decrease from 40 bp. Thus a combinatorial screening approach is best used when oligos are 35-40 bp or longer.
    - A potentially less computationally intensive approach would be to conduct a “pre-screen” of oligo pools that looks only for some threshold of sequence overlap (e.g., 8, 10, or 15 bp) between a positive strand of one oligonucleotide and negative strand of a second. For those overlapping oligos, an algorithm could next calculate their Tm. If that preliminary screening identified overlapping oligos with similar Tms, that would be evidence of intent to assemble and would trigger further combinatorial screening. With this approach, the length of the oligos matters less than the length of the overlap; screening 20mers might be feasible with 15bp overlaps. However, assembly plans do not universally require uniform Tms, so prescreening on that basis might not be sufficient.
- The 1 µmol quantity standard is unlikely to meaningfully increase security. Assembly is routinely accomplished with quantities below 1 µmol. Additionally, the differing capabilities of industry to synthesize pools at very low quantities (e.g., femtomoles vs nanomoles) would result in some organizations essentially needing to combinatorially screen all oligo pools while others rarely would do so.

Best match or equivalent: standards and metrics

Proposed Revised Guidance notes that providers may choose to use screening approaches they deem equivalent or superior to Best Match. This recognition that alternative screening systems may be equal to or superior to Best Match is welcome. However, it highlights a missing element from the Guidance and screening ecosystem. *There are no standards that define Best Match performance, and therefore no metrics that can measure an alternative approach against Best Match*. Workshop participants indicated that standards and metrics would provide significant value to providers and meaningfully contribute to security. One EBRC workshop participant noted prior research that indicated that some DNA synthesis companies believed they were in compliance with Guidance, but in fact fell short. Another workshop participant pointed to the Genome in a Bottle Consortium hosted by NIST as a potential model for bringing private and public sector experts together to develop standards and a conformity assessment mechanism. NIST has the expertise to support the development of such, but would need supporting, on-going resources. Alternatively, an existing non-profit, public-private partnership could be identified to lead this effort in the U.S. On a global scale, it could be beneficial for an international organization with a biosecurity and biosafety focus (e.g. International Biosafety and Biosecurity Initiative for Science (IBBIS), soon to be launched by NTI) to be involved or lead the development of standards and metrics to increase international adoption. It was further noted that the utility of currently available databases for Best Match are not fit for purpose; many records are misannotated or contain sequences of genetic tools.

Metrics could incorporate rates of false positive and false negative findings against test sequence sets and evaluate capabilities for detecting obfuscated sequences. Metrics and evaluation standards would need to be regularly updated in response to evolving capabilities and concerns. The development of standards and metrics would necessitate some degree of biological red-teaming, pressure testing, and/or auditing ; nefarious actors may use sequence obfuscation or other methods to out-maneuver sequence screening methods. Metrics should be able to evaluate a system's capacity for identifying obfuscated sequences. Concerns about information hazards can emerge from these practices, but these can be minimized and contained. There is value in good actors identifying and attending to vulnerabilities before they can be taken advantage of by nefarious actors.

4: The Screening Ecosystem

Boosting Guidance compliance

- Boosting compliance with carrots: Many have suggested boosting compliance by requiring public funds *e.g*. from NIH, NSF, etc be used to buy DNA only from Guidance-compliant providers. This would be a powerful incentive, though would have some implementation challenges:
    - Some organization or entity would need to certify compliance so that customers knew which providers to use. This would require standards and metrics (covered in Part 1 of EBRC response). Alternatively, membership in IGSC could be sufficient, however 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.
    - Individual institutions would need to know which providers were Guidance-compliant and ensure that orders were placed only with those providers. Current practices for buying oligos at academic institutions would make this challenging—any solution where

individual oligo orders pass through procurement would create significant slow downs that researchers would deem unacceptable.
    - Customer education would be needed.
- Boosting compliance by removing obstacles: Providers might choose not to screen out of concern for disruptions to operations and the cost of building, implementing, and maintaining screening systems. Screening as a service by a third-party or the availability of readily adoptable screening platforms could diminish these obstacles.
- Boosting compliance with sticks:
    - It was suggested that, because Select Agent and Export Administration Regulations have the force of law, USG could require providers who make gene-length products to demonstrate that they are not synthesizing and transferring/exporting regulated materials.
    - Companies without screening could be required to purchase liability insurance. This would require some measurement of screening performance and should only be considered in conjunction with the availability of resources to minimize the burden of implementing minimal (floor) screening.

International considerations

- As evidenced by COVID-19, biological threats are not contained by international boundaries. Joint security exercises can build international commitment and competence. Funding for such initiatives is needed.
- International screening brings up many questions including who decides what needs to be screened for and who has access to that information. There was general support for an international body to bring international stakeholders together to collaboratively build norms and best practices for a secure screening ecosystem.
- Research and development in the life sciences is an international endeavor. Every choice or element of security measures within the United States should be considered within that broader context. The US can continue to lead, while collaborating/discussing these important issues with international counterparts. This can help ensure that the steps taken by USG are useful beyond our own borders to make everyone safer, more secure, and able to reap the benefits of a robust R&D ecosystem.

Follow-up screening

Follow-up screening is challenging for providers as they attempt to suss out whether or not their customer might (deliberately or unintentionally) cause harm with the ordered sequences. These are important decisions that have to be made relatively quickly and with limited information. Additional training, specific best practices, example scenarios, etc would be of great use to industry. As examples, a customer might provide the name of a biosafety officer at their company, but how does a provider verify the credentials and trustworthiness of that person? How can or should a provider verify documentation of a start-up located outside the US? Should concerns be “alleviated” if the customer withdraws their order? What if they state an intent to order from a different provider due to frustration about the screening process? What standard should be met for contacting an FBI WMD Coordinator?

Some of these follow-up interactions might be smoothed by USG providing sample questions or statements to use with customers during follow-up screening. This would increase consistency across industry and give industry some top cover.

As USG considers how it will define SOCs, it is appropriate to continually recognize that "concern" is not binary. Some SOCs are *very* concerning, others less so. As database tiers are considered, it might also be appropriate to consider follow-up tiers. For example, providers could follow-up immediately on any sequences in the most concerning tier. Sequences in "yellow" tiers could be flagged, but not triggered for follow-up unless associated sequences are also ordered. Or, if a given customer ordered some number of yellow flag orders in a given time frame, that could trigger follow-up. Here again, though, clear guidance on what constitutes and alleviates concerns would be needed.

5: Benchtop Synthesis Equipment

The inclusion of manufacturers of benchtop synthesis equipment is an important addition to the proposed Revised Guidance. Benchtop synthesis has a unique suite of challenges, including whether sequence screening should always "phone home" to the manufacturer or if local screening on a synthesis device can provide sufficient security.

Benchtop customer screening

The emphasis on robust customer screening by manufacturers of benchtop equipment is appropriate. A few points of clarification would enhance the usefulness of the Revised Guidance for manufacturers.

- Customer legitimacy: Verifying legitimacy of a customer of a benchtop device might involve different evidence or standards than verifying legitimacy for purchasing a SOC.
- Equipment "appropriate for [customer] needs:" What criteria or documentation demonstrate customer needs? If a customer wants to buy a device that has advanced capabilities that the manufacturer thinks the customer does not need and might not use, should a manufacturer encourage (or insist upon) an alternative model?
- "Manufacturers and their Customers should implement mechanisms to track continuously the legitimacy of users of their equipment:" It is unclear if the manufacturer or the customer is ultimately responsible for this. It is also unclear if this statement refers to *all* equipment users. For example, if a new graduate student, scientist, or research assistant joins a laboratory that operates a benchtop synthesis machine, should they be tracked if they might occasionally use the synthesis device? If the manufacturer has the role of tracking legitimacy of users, does the customer have an obligation to report the legitimacy of new lab members to the manufacturer? Or should the lab provide the name of the user to the manufacturer, who would screen it? It might be more appropriate for manufacturers to keep records of the legitimacy of Primary Users instead of all users as this would become a significant administrative burden that might not add significant security. Another option that would track use more than legitimacy is for users to authenticate before every run of the machine and for that data to be kept for eight years.

The draft Revised Guidance also says that "If the customer indicates plans to produce SOCs," the manufacturer should conduct prescreening. This language suggests that the manufacturer does not have a responsibility to ask customers if they intend to synthesize SOCs. If not directly asked, customers might not be forthcoming about potential SOC synthesis, especially if they do not have immediate plans to

produce SOCs. And, current synthesis needs are not always indicative of future needs and projects. Thus, hinging prescreening on customer indication of a plan to produce SOCs might result in such prescreening being inconsistently applied, and more likely to be applied to customers who are already security conscious. Additionally, it is not clear what is meant by "prescreening mechanisms to determine legitimate use." Therefore, uniform, rigorous customer screening, independent of stated equipment use intentions, might be most appropriate.

Benchtop sequence screening

There was significant discussion at the EBRC workshop about whether or not sequence screening locally on a synthesis device could provide sufficient security, or if a "phone home" approach is necessary. Phone home approaches enable manufacturers to flag and potentially halt the synthesis of SOCs and/or scan for unusual patterns of synthesis. Screening algorithms and databases can be updated in response to emerging or diminishing concerns. However, some customers cannot, will not, or prefer not to purchase internet connected devices, so if Guidance-compliant manufacturers do not offer devices with local screening capabilities, customers might turn to non-compliant manufacturers that lack screening capabilities all together. Customers may avoid equipment that phones home out for various reasons, such as IP and cybersecurity concerns. Cybersecurity approaches that encrypt sequence information could be implemented, but customers would have to trust those approaches. Other customers around the world, e.g. at remote research stations or non-urban areas, might not have consistent high-speed internet access.

Local screening systems, however, are generally less secure. They could be hacked or disabled without the manufacturer becoming aware. Updating a local screening system in response to emerging or diminishing concerns would be very challenging without internet connectivity, expensive site visits from manufacturers, and/or shipment of a drive with update capabilities.

Revised Guidance could strongly encourage phone home approaches, but recognize the market pressure companies face to offer local screening mechanisms. It could describe best practices that enhance the security of locally screening devices when a phone home device is not an option, for examples:

- Manufacturers could ensure that organizations that purchase locally screening equipment have a biosafety and/or biosecurity officer who meets with the manufacturer and Principal User in advance of purchase and at intervals (e.g., annually, biannually) thereafter to discuss equipment use and anticipated SOC synthesis. The extent to which such a biosafety/biosecurity officer is responsible for overseeing the use of the device has significant implications that should be carefully considered. Institutional biosafety committees are trusted to oversee broad swaths of research that carry significant biosafety and biosecurity risks. However, it is unreasonable to expect such officials to have oversight of all sequences synthesized within their institutions. One potential security increasing measure is to require biosafety/biosecurity officers to enter a unique approval code when a SOC is identified via local screening. Additionally, biosafety/biosecurity officers at institutions around the world may have different policies and responsibilities. Verifying the qualifications and relevant role of a biosafety or security officer might pose a significant challenge.
- Manufacturers could charge customers a fee for periodic site visits by the manufacturer to run analysis on the screening system and its records and to install software updates.

- Locally screening equipment could require user authentication for each sequence synthesized. (This might also be appropriate for equipment that phones home.)

Similar to providers conducting follow-up screening, manufacturers would benefit greatly from having clear language in the Revised Guidance that they might share directly with customers to explain screening and security processes.

Security over equipment lifetime

The emphasis in the proposed Revised Guidance on tracking equipment through its lifecycle is very important. Closed loop systems, in which equipment is bought back by companies for refurbishment and resale would decrease the possibility of a customer reselling on a secondary market without customer screening capacity. Alternatively, manufacturers could stipulate in purchase agreements that potential resales must be reported to them for new customer screening. The temptation to circumvent this safeguard could be dampened by limiting the sale of proprietary reagents to registered customers. It may be prudent to make customers aware of relevant export control regulations to communicate the importance of adhering to resale procedures.

6: Responsiveness to Emerging Technology

Biotechnologies are rapidly advancing that will present new obstacles to synthetic oligonucleotide synthesis security. Research that can strengthen and support the Guidance is also being undertaken. Thus, USG should institute a structure that allows for regular review and, as necessary, update to the proposed Revised Guidance.

Technologies that might challenge Guidance

- Recoding the genetic code: Researchers are developing orthogonal translation pathways that utilize non-canonical amino acids and/or other monomers to build non-canonical biopolymers. In some cases, just a few recoding events could avoid BLAST detection, but more frequently, a significant number of codons need to be reprogrammed to avoid detection by screening algorithms. Additionally, researchers are improving upon techniques to incorporate unnatural base pairs into DNA and transcribe it to RNA. These active areas of research will advance rapidly in the coming decade. Therefore, USG could support research to explore how/if/when these developments might challenge sequence screening and possible solutions.
- Whole genome design and editing: The capability to make many very specific genomic edits and modifications simultaneously allows researchers to tap into the depths of genetic possibility and discovery. This technology has many use cases, such as engineering an organism to produce higher titers of a desired product through engineering of regulatory sequences, controlling gene expression, and optimizing protein function and pathways. These outcomes depend upon genomic context. As platforms for genome design and editing are currently available, Guidance should work to address best practices for these companies and customers.

Technologies that might strengthen Guidance

Several emerging technologies might also be able to strengthen the ability of providers, customers, and other stakeholders to enact best security practices for synthetic DNA. One approach to help address concerns such as reagent switching on benchtop devices, sequence obfuscation, and genetic recoding is a pattern-based approach to sequence screening (https://doi.org/10.1109/TKDE.2015.2510010). Another

important approach is functional annotation and prediction. AI/ML advances are enabling greater predictive capabilities for novel protein development. Those algorithms could be applied to ordered sequences without significant homology to known sequences to help identify novel SOCs. Finally, the importance of standards and metrics was described previously, but is noted here as an area that needs research and development support.

A concluding thought

Liability was an undercurrent that ran through much of the workshop, occasionally surfacing to direct discussion. It, of necessity, plays a role in shaping how different stakeholders think about screening, and the impact of those concerns should be acknowledged and taken into account. While the motivations of all stakeholders are complex and multi-dimensional, they ultimately stem from a commitment to the responsible advancement of life sciences research and development; there is obvious room for common ground and compromise.

Industry might advocate for greater specificity in the Guidance so that liability for product misuse can be better assessed based on demonstrated compliance. Guidance that clearly articulates the role of each stakeholder helps all understand and protect their interests into the future, and provides a backdrop for rational discourse where the government is a valued partner. If the Guidance is vague and open to interpretation, the burden of uncertainty and risk is shifted to the future economic and legal landscape. The vastness of the threat space and the tradeoffs between iron-clad security and progress of research might deter USG from providing the degree of clarity desired by industry.

To some extent, avoiding Guidance overreach would be useful for encouraging all parties to comply. On the other hand, fear of overreach should not prevent stakeholders from identifying and implementing screening that will have the greatest benefit without slowing the ground-breaking research and development in the biosciences, based on a community model (as seen herein) where the government is a partner. Funding to explore liability possibilities and frameworks, potentially through table-top exercises and reports could be useful.