The Four Most-Common Organizational Challenges in Successful Multiplex Panel Design

Introduction

In this white paper, we will discuss the four most frequently encountered problems an organization faces in multiplex PCR panel design. These include ***development time, development cost, expertise limitations,*** and the use of ***suboptimal tools***. Developing multiplex panels is a complex and challenging endeavor. Most molecular diagnostics firms consider assay development to be a core competency of the organization and often will solely rely on the resources available within the company. This can result in a variety of undesirable outcomes including costly trial-and-error experimental iteration, lost opportunity in delays bringing the diagnostic to market, and in many cases the outright inability to design an efficacious multiplex assay.

Development Time...Impacts Time to Market

Primer design is a critical early-stage step in multiplex PCR panel development. The difficulty to successfully design multiplex primers and probes is often underestimated and the complexity of the problem is many times undervalued. Most, if not all, of the downstream assay challenges the organization will later encounter, including lost time and money, are the direct byproduct of the lack of proper design at the inception.

Depending on the complexity of the panel, trial-and-error approaches can take an average of six months to one year and in many cases well over a year when including iteration and optimization processes. The process typically involves multiple employees (3 – 4 FTEs) working on not only the *in-silico design* but also the wet laboratory empirical testing and validation. If the initial designs fail, as they often do, the trial-and-error cycle repeats itself until an acceptable end solution is determined. Based on aggregate data compiled through DNAS customers, we have found these costs to average approximately $300,000 per multiplex reaction when accounting for employee salaries, reagents and related resources, including overhead. This estimate does not include the lost opportunity cost for the inability to bring products to market, which likely exceeds the direct costs of failed assay design.

> "When there are multiple reactions working simultaneously the complexity increases exponentially. It gets to a point where it is impossible for a human being to rationally figure it out in any logical or realistic way. You can get lucky, or use trial and error but you are just fooling yourself."
>
> Arjang Hassibi, Ph.D., CEO, InSilixa

Figure 1 illustrates the assay design paradigm frequently utilized by most organizations. The perception of "free" software can be very misleading as this design cycle will create an expensive infinite loop of iteration, experimental testing and redesign. The consequences of repeating this process are time delays that can disrupt mission-critical outcomes such as requisite diagnostic approvals and the advantages that come with being first to market for a given diagnostic or platform.

Figure 1: The trial-and-error design cycle of failure.

Even when "successful", the designs from experimental iteration are suboptimal from the perspectives of sensitivity, specificity, and robustness to varying reagent lots and sample contamination. Clearly, there is a need for an improved approach to multiplex design.

> "The challenge is having the detection of the intended target versus detection of nonintended targets; making sure every time you detect the target you want but only the target you want."
>
> Aude Argillier, Senior Scientist, Design and MDx Assay Development, QIAGEN, Manchester, UK

Thus, even after the assay is introduced, the trial-and-error design cycle may begin again if broader market usage uncovers false negatives or false positives for the initial design that mandate test revision or even restarting an assay design from scratch. If PCR assay design is critical to a company's mission, then more efficient approaches are needed to accelerate the discovery work that is essential to that company's success.

Development Cost...Computational versus Trial-and-Error

Time and effort for trial-and-error approaches have significant associated costs. Not only are there fully-burdened employee costs, but also associated laboratory and reagent costs in addition to lost revenue due to delay in market-entry (Table 1). In a highly-competitive market, product introduction delays can mean the difference between securing market leadership as opposed to being second, third or an "also ran" product offering.

Development Cost
3 – 5 Fully-burdened employees
Access to computational power
Laboratory space, reagents and supplies
Lost revenues due to delayed market entry
> $300,000 per multiplex assay

Table 1: "The Cost of Free".

> "A lot of time and effort goes into designing assays yet a lot are on version 2 or 3. Sometimes after a product is introduced you start seeing false negatives and portions of the assay require rework."
>
> Jaime Prout, Developmental Scientist II, Beckman Coulter

Most researchers resort to the empirical iterative approach because they don't recognize that there is an alternative. Such an empirical approach often starts with optimization of individual singleplexes. These singleplexes are then combined into larger and larger multiplexes until a failure is identified (i.e. a member of the multiplex does not amplify efficiently or false amplicons are produced). The members of the multiplex are then changed (without knowing the actual reason why they failed) and the modified multiplex is then tested again. This "linear 1-dimensional" search does not work. The reason why this fails is that Multiplex PCR is NOT a linear system! Instead, multiplex PCR is a non-linear multidimensional landscape with complex interactions among the variables. It is not unusual to have a 7-plex PCR assay working only to have it completely fail when an eighth primer pair is added. Efforts to make higher multiplexes often completely fail even with highly qualified teams expending large resources. The process is much like the arcade game "Whack-a-mole" where you knock down one problem just to have another problem pop up. This is practically the definition of insanity!

Expertise Limitations…Too Many Details

Scientists designing PCR assays are usually subject-matter experts in molecular biology and/or bioinformatics. However, such design teams often lack expertise in the biophysics of nucleic acid thermodynamics, chemical kinetics, specialized algorithms for predicting hybridization, detailed knowledge of enzyme characteristics, and advanced algorithms for optimization. Most companies lack these resources and personnel; they do not have this type of expertise on staff. As a result, a knowledge gap exists and scientists that typically are formally trained in molecular biology or bioinformatics find themselves taking a crash course in biophysical chemistry and thermodynamics and optimization algorithms to better understand why their assay keeps failing.

Most assay-design groups tend to select primers based upon primitive Tm predictions and naïve computer programs. They may consider crude models for hairpin formation in primers or formation of primer dimers. Rarely, however, would they consider the crucial effect of competing target secondary structure, which is the dominant cause false negatives. They also may underestimate the false positives that can result from phylogenetically similar organisms or other background organisms such as the human genome. Sequence variations are also an important source of false negatives, and properly dealing with such variation to maximize coverage is quite challenging.

Lastly, very few assay design groups will account for all the interacting parts of a multiplex reaction and instead try to optimize the multiplex reaction with a linear 1-dimensional approach. However, multiplex PCR is a nonlinear complex system with many interacting variables, and there are myriad reasons for failure. In multiplex PCR, there are many targets and thousands of primer candidates, which results in a combinatorial explosion in the number of cross-hybridization interactions and false amplicons. It is impossible for a human to account for all such interactions. In addition, if the singleplex reactions are not designed correctly, then unequal amplification rates can also affect the dynamics of the multiplex reaction (that can be minimized using the principles from section 1). Finding the combinations of primers that do not interfere with one another is an optimization problem with a multidimensional landscape with a huge number of possibilities.

> "Relying more on computer-aided designs instead of empirical work drastically reduces labor and reagent costs. We have proven that we can have 70% of the job done with the computational design and the remaining 30% is empirical effort. That definitely reduces our costs."
>
> Arjang Hassibi, Ph.D., CEO, InSilixa.

Computer algorithms that incorporate the biophysical or thermodynamic expertise to finesse model optimization would be of great benefit, and allow staff to focus their expertise on the subject matter they know best that is crucial to the assay design.

Use of Suboptimal Tools… The Cost of Free

Many available freeware tools address certain aspects of multiplex panel design, and many researchers cobble together multiple freeware applications as a customized primer design pipeline. However, most freeware was not developed for the complexity of multiplex PCR panels. The limitations of freeware are further exacerbated when researchers try to piece together singleplexes into a multiplex reaction. The cost of free can be iterative cycles of development and testing (Figure 1).

Common mistakes made by assay design teams is the use of freeware such as *BLAST* (basic local alignment search tool), to detect cross-hybridization. The inappropriate use of *BLAST* contributes to a number of assay design problems. For example, *BLAST* provides too many irrelevant hits, the wrong ranking of hits, misses about 80% of thermodynamically-stable hits, does not distinguish extensible from non-extensible hits and it does not detect amplicons. These failures happen because *BLAST* is meant to determine sequence similarity to infer common evolutionary ancestry and thereby infer function. However, sequence similarity does not equal thermodynamic stability of complementary sequences. A tool that natively scores oligos based upon the thermodynamics of hybridization is *ThermoBLAST* from DNA Software, Inc., which is described in detail in the following webinar:

https://www.youtube.com/watch?v=rrClaqNYYOA&t=14s

Many users rely on Tm prediction tools that are based on primitive models of hybridization and neglect the effect of competing secondary structure.

> "Combining tools, such as MFOLD, Primer3, and Oligo ANALYZER, in just the right way may produce some interesting designs that work fairly well, although they are frequently limited to 3-5 multiplexes. It would be simpler and more time effective to have an integrated application for multiplex design."
>
> Andrew Dunn, Senior Scientist, Preclinical and Clinical Assays, CRISPR Therapeutics

Primer3 uses proper thermodynamic parameters, but it is limited to the 2-state model that completely neglects the competing effects of secondary structure in both the oligonucleotides and, even more important, the target. *Primer3* also does not properly account for the effect of dangling-end nucleotides and does not account for the very important effect of magnesium concentration. *MFOLD* is a wonderful tool for predicting unimolecular folding, but it does not account for competition between unimolecular and bimolecular reactions and target refolding that occurs upon hybridization. The proper way to account for competing structure is to use the "multi-state coupled equilibrium model" (SantaLucia, J., Jr. and Hicks, D. "The Thermodynamics of DNA Structural Motifs," *Annu. Rev. Biophys. Biomol. Struct.* 33, 415-40 (2004).). This more advanced model is included in the *"Oligonucleotide Modelling Platform", OMP,* from DNA Software, Inc.; notably, *OMP* is directly incorporated into *PanelPlex* (discussed below). The *"Oligo ANALYZER"* (Integrated DNA Technologies) has crude folding algorithms and is not set up for large multiplex reactions.

Lastly, none of the free software solutions take into account collections of whole genome sequences. Genome sequencing projects have produced a treasure trove of information on pathogen and human variation and normal background flora. *ThermoBLAST* was specifically created to address this need. For example, designing a diagnostic for the 2009 pandemic H1N1 influenza A, should account for all the >7000 genome variants that have been sequenced to date (this is the inclusivity list). In addition, such a diagnostic should not give false positives to the human genome, human transcriptome, or near-neighbor (i.e. phylogenetically related) viruses such as influenza B or H3 variants of influenza A (these are the background and exclusivity lists). Importantly, these genome database searches require massive computational resources, that most organizations lack. Fortunately, *ThermoBLAST* is implemented using cloud computing so that any group that licenses that product now has the power to utilize the full resources of modern genomics. Moreover, *ThermoBLAST* is directly integrated into *PanelPlex* (described below) so that it can account for sequence variations in the inclusivity panels, and the off-target hybridizations that occur to near-neighbor and background organisms.

What Would Solve These Problems and Provide for Better PCR Primer Design?

To resolve the complexity of consensus and multiplex panel design, many factors must be incorporated into the ideal solution. The appropriate algorithms and models must include the correct thermodynamic scoring, weighting and filtering of candidates while using the proper methods to check for off-target effects and unintended hybridization. This solution cannot be synthesized through combining multiple freeware packages and the magnitude of the problem is too great to overcome through brute force or experimental trial and error. This multifactorial problem can only be solved through leveraging large-scale cloud computing in combination with the right predictive models and wet-lab validated algorithms.

A Pragmatic Solution… *PanelPlex*

PanelPlex provides completely automated design of multiplex PCR with unprecedented sensitivity, specificity, and coverage. *PanelPlex* is the culmination of more than 15 years of investigation into the mechanism of PCR and careful experimentation to identify the sources of PCR failure. *PanelPlex* has been rigorously validated for the detection of numerous viruses, bacteria and human targets. The current version of *PanelPlex* is focused on consensus design for applications such as: infectious disease variants for bacteria, viruses, human genomic targets, and mRNA profiling. A new module called *"MultiPick"*, has been integrated into *PanelPlex* to create a new product called *"PanelPlex-NGS"* that allows for multiple assays to be combined into large multiplexes. This algorithm is useful for designing the oligos for applications such as target enrichment for next-generation sequencing (NGS) and highly-multiplexed molecular diagnostics with exquisite sensitivity and specificity. In conclusion, *PanelPlex* runs billions of computations and exhaustively evaluates solutions to eliminate the iterative trial-and-error design cycle paradigm, thereby greatly reducing wet-lab optimization and thereby greatly reducing developmental time and cost.

Figure 2 summarizes the simple user workflow for *PanelPlex*. Step 1 is to input all the target accessions for the inclusivity, exclusivity, and background lists. Step 2 is to input the annealing temperature and master mix composition. Step 3 is to input advanced parameters such as the fluorophores used on the probes and the weighting parameters for each scoring term. *PanelPlex* then automatically produces the designs.

More details regarding the features, user input, and best practices are given in the document "*PanelPlex* User Guide". More details about the algorithms and scoring used by *PanelPlex* are given in the Whitepaper entitled "The Four Most Commonly Encountered Problems in Multiplex Panel Design". Both of these documents are available upon request at the website: www.dnasoftware.com.

Figure 2: *Simple user workflow in PanelPlex.*

Summary

PanelPlex solves the four most frequently encountered organizational problems in multiplex panel design: development time, development cost, expertise limitations, and the use of suboptimal tools. PanelPlex is a next-generation cloud-based application that represents the best of what DNA Software is known for: difficult targets, high-level multiplex, assay optimization and the elimination of trial and error approaches to help companies more efficiently build better diagnostics.

PanelPlex provides:

- Quick Results – Even complex multiplex designs can be completed within 24 hours
- Expertise – Easy access to “best in class” algorithms, models, computational power and thermodynamics expertise
- Industry Validation – wet-lab validated results, commercial grade solution
- Economical – Is a fraction of the cost of what organizations spend on “traditional” assay design
- One-Stop Shopping – Eliminates the use of suboptimal tools and inappropriate resources

PanelPlex is a trademark of DNA Software, Inc.
©2017 DNA Software, Inc. All rights reserved.

334 E. Washington St. | Ann Arbor, MI 48104 | 734.222.9080
www.dnasoftware.com | info@dnasoftware.com