ROUTING AND ACTION

MEMORANDUM

ROUTING

TO:(1) Biological & Biotechnology Sciences Branch (BBS) (Strand, Micheline)

Report is available for review

(2) Proposal Files Report No.: Proposal Number: 67518-BB.1

DESCRIPTION OF MATERIAL

CONTRACT OR GRANT NUMBER: W911NF-16-1-0022

INSTITUTION: University of California - Berkeley

PRINCIPAL INVESTIGATOR: Jasper Rine

TYPE REPORT: Final Report

DATE RECEIVED: 11/8/23 11:53AM

PERIOD COVERED: 12/29/15 12:00AM through 3/25/23 12:00AM

TITLE: Final Report: Actionable Variation in Human Genes to Enhance Metabolic Efficiency

ACTION TAKEN BY DIVISION

(x) Report has been reviewed for technical sufficiency and IS [x] IS NOT [ ] satisfactory.

( ) Based on my technical review, I have identified no OPSEC or Technology Protection concerns that need to be addressed regarding this report.

(x) Perfomance of the research effort was accomplished in a satisfactory manner and all other technical requirements have been fulfilled.

(x) Based upon my knowledge of the research project, I agree with the patent information disclosed.

Approved by SSL\MICHELINE.STRAND on 11/9/23 11:08AM

ARO FORM 36-E

REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
08-11-2023	Final Report	29-Dec-2015 - 25-Mar-2023

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Final Report: Actionable Variation in Human Genes to Enhance Metabolic Efficiency	W911NF-16-1-0022
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102 611103
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of California - Berkeley The Regents of the University of California 1608 Fourth Street, Suite 220 Berkeley, CA 94710 -1749	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 67518-BB.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jasper Rine
UU	UU	UU	UU		19b. TELEPHONE NUMBER 510-207-8465

RPPR Final Report
as of 21-Dec-2023

Agency Code: 21XD

Proposal Number: 67518BB
Agreement Number: W911NF-16-1-0022

INVESTIGATOR(S):

Name: Jasper Donald Rine
Email: jrine@berkeley.edu
Phone Number: 5102078465
Principal: Y

Organization: **University of California - Berkeley**
Address: The Regents of the University of California, Berkeley, CA 947101749
Country: USA
DUNS Number: 124726725
EIN: 946002123

Report Date: 25-Jun-2023
Date Received: 08-Nov-2023
Final Report for Period Beginning 29-Dec-2015 and Ending 25-Mar-2023
Title: Actionable Variation in Human Genes to Enhance Metabolic Efficiency
Begin Performance Period: 29-Dec-2015
End Performance Period: 25-Mar-2023
Report Term: 0-Other

Submitted By: Jasper Rine
Email: jrine@berkeley.edu
Phone: (510) 207-8465

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 0
STEM Participants:

Major Goals: MAJOR GOALS

Nutritional vitamin deficiencies result in severe disorders such as anemia, cardiovascular defects, neurological malfunction, and atrophy (among others). At the cellular level, many metabolic pathways are compromised by nutritional deficiencies such as: increased DNA damage (folate, B12, niacin deficiency), decreased neuronal connectivity (thiamine, B6 deficiency), redox imbalances which lead to oxidative stress (riboflavin deficiency). Some of these gross phenotypes are due to dietary inadequacies and studies related to this contributed to establishing the Recommended Daily Allowance (RDA), which approximates the amount of each vitamin and mineral needed to prevent loss of health for the average person under an average range of experiences.

However, some metabolic imbalance is due to single enzyme deficiencies (as is the case for MTHFR above). When the RDAs were established in the middle of the 20th century it was not possible to describe or incorporate individual genetic variation in the recommendations. From our initial studies, we now know that certain nonsynonymous substitutions in enzymes require significantly higher levels of their vitamin cofactors for normal performance. Hence the RDA is inadequate to the needs of some individuals, especially warfighters and athletes, who are asked to perform at the upper extremes of human capacity.

Objectives of the Study:

Our first objective is to bring greater value to the growing catalog of human genetic variation by identifying which of these variants result in suboptimum performance of individual enzymes. The second objective is to determine which of the suboptimum variations are amenable to metabolic tuning and which are not. This will be accomplished by leveraging the resources and lessons learned from our previous work with folate metabolism and vitamin B6 utilization. Our goal is to extend those results to a greater number of enzymes amenable to this strategy of functional remediation. In this proposal, we are focusing on nutritional optimization of metabolic steps essential for generating cellular energy stores (ATP production and mitochondrial function) as this would have the greatest impact on enhancing warfighter performance (physical and cognitive), especially under stressful conditions.

Scope of the Proposed Study:

This project is designed to expand our work with folate and vitamin B6 utilizing enzymes to other important cofactors and pathways in a way that could have the greatest impact on peak performance and training. Consequently, this study will focus on the potential for nutritional tuning of metabolic steps essential for generating cellular energy stores (ATP production and mitochondrial function). In addition to the obvious effect of these reactions on physical output, neural activity and brain information processing depend critically on an adequate energy supply. Moreover, this study focuses on variation in the genes that produce the activated cofactors, which

would be most likely to have synthetic effects in combination with hypomorphic alleles of genes encoding enzymes that use these cofactors. Thus, these collective reactions have a considerable influence on physical and cognitive performance.

Bioenergetics, or cellular respiration, is meant here to be the amalgam of reactions responsible for carbohydrate breakdown and electron transfer in the presence of oxygen to produce ATP. This metabolic cassette is often discussed as comprising 3 functional compartments: glycolysis, the TCA cycle and oxidative phosphorylation (the latter 2 reside within the mitochondrion). Within these pathways, vitamins and cofactors play a critical role. For example, a major electron donor/acceptor is NAD (nicotinamide adenine dinucleotide) which is synthesized from vitamin B3 (niacin). In addition, flavin cofactors (FAD, FMN) synthesized from riboflavin are necessary for several enzymatic steps (e.g., succinate dehydrogenase), as is the cofactor thiamine pyrophosphate (e.g., pyruvate dehydrogenase), which is synthesized directly from thiamine.

It is clear, therefore, that the flux through this metabolic cassette is dependent on the utilization and modification of these 3 vitamins into their respective cofactors/coenzymes. We postulate that inefficiencies in this utilization (due to impaired enzyme variants) may compromise subsequent energetic reactions in a way that may be remediated, or optimized, by supplementation of the cognate vitamin. Thus, we choose to focus on 5 key enzymes involved in cofactor/coenzyme synthesis. These are described:

1. RFK (Riboflavin kinase). This enzyme catalyzes the phosphorylation of riboflavin (vitamin B2) to form flavin mononucleotide (FMN), an obligatory step in vitamin B2 utilization and flavin cofactor synthesis.
2. FLAD1 (FAD synthase). Performs the second step in the synthesis of the cofactor flavin adenine dinucleotide (FAD) from riboflavin.
3. TPK1 (Thiamine phosphokinase). This enzyme phosphorylates thiamine (vitamin B1) to produce the cofactor thiamine pyrophosphate, an obligatory step in vitamin B1 utilization.
4. NAPRT (Nicotinate phosphoribosyltransferase). Acts in the salvage pathway of NAD+ biosynthesis. Performs the first modification of vitamin B3 (niacin/nicotinic acid) to form nicotinic acid mononucleotide, an obligatory step in vitamin B3 utilization.
5. KYNU (Kynureninase). This enzyme catalyzes the 3rd step in the de novo biosynthesis of NAD+ from tryptophan.

For each target enzyme, we intend to interrogate most, or all described nonsynonymous SNPs for functional impairment, metabolic impact and vitamin-remediability in the assays described below. We will determine which changes are benign and which have a functional consequence. Further, we will determine which may respond to nutritional tuning via cofactor remediation. These studies will build on our previous datasets to further understand the characteristics of remedial alleles and to develop better predictive factors from genome variation data.

Accomplishments: see PDF upload

Training Opportunities: Nothing to Report

Results Dissemination: Nothing to Report

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Co-Investigator
Participant: Nicholas Marini
Person Months Worked: 12.00
Project Contribution:
National Academy Member: N

Funding Support:

Participant Type: Technician

RPPR Final Report
as of 21-Dec-2023

Participant: Jean Yan
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Jasper Rine
Signature Date: 11/8/23 11:53AM

ACCOMPLISHED

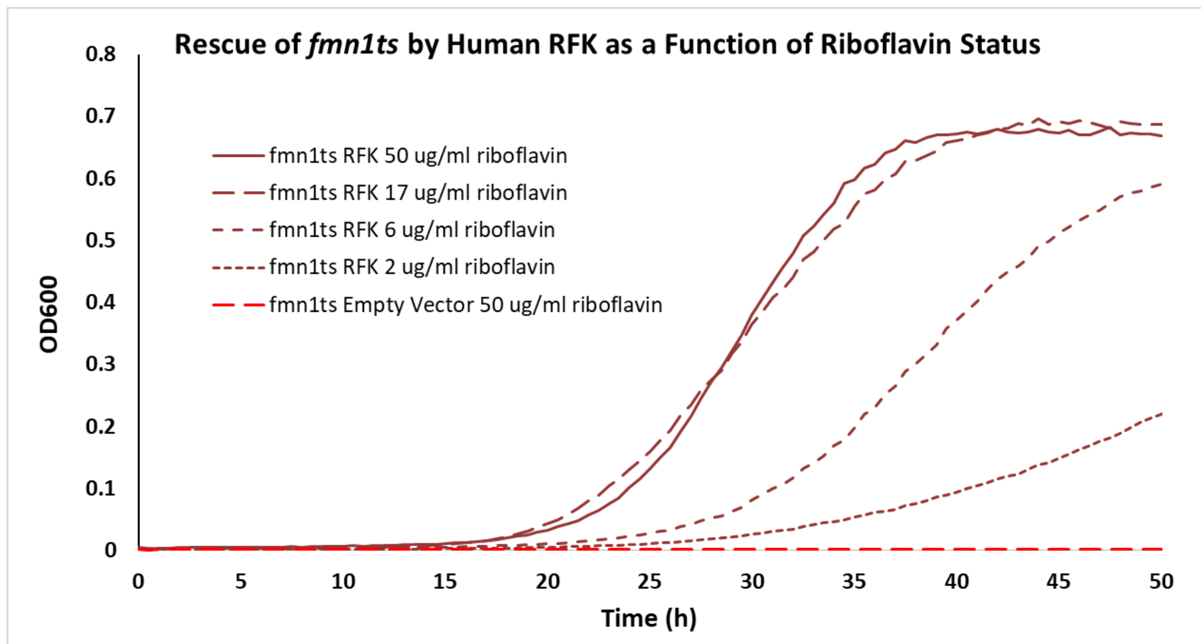
This project is designed to understand the utility of coupling cofactor supplementation to personal genetic information in a way that could have the greatest impact on peak performance and training. Consequently, this study will focus on the potential for nutritional tuning of metabolic steps essential for generating cellular energy stores (ATP production and mitochondrial function), which have a significant influence on physical and cognitive performance. As detailed in the Major Goals section, we have chosen to focus on 5 key enzymes involved in these processes that may be amenable to cofactor augmentation and for which common genetic variants exist that may affect enzyme function: 1) *RFK* (Riboflavin kinase), 2) *FLAD1* (FAD synthase), 3) *TPK1* (Thiamine phosphokinase). 4) *NAPRT* (Nicotinate phosphoribosyl transferase), and 5) *KYNU* (Kynureninase). For each target enzyme, we intend to interrogate all described nonsynonymous SNPs for functional impairment, metabolic impact and vitamin-remediability in functional assays. We will determine which changes are benign and which have a functional consequence. Further, we will determine which may respond to nutritional tuning via cofactor remediation. These studies will build on our previous datasets to further understand the characteristics of remedial alleles and to develop better predictive factors from genome variation data.

Initial-Phase Functional Assays: Evaluating Nonsynonymous Variants for Metabolic Impact and Nutritional Remediation in Quantitative Yeast Complementation Assays.

In this approach, the *in vivo* function of human target enzymes or (enzyme variants) are measured individually in a cell-based assay in which the human gene is expressed and functionally complements a yeast cell that has had the yeast gene for the orthologous enzyme removed from the chromosome. Thus, the rate of growth of yeast cells reflects the relative activity of each nonsynonymous variant of that human enzyme. This strategy has several desirable qualities; (1) expression of the human clone can be driven by a heterologous promoter and terminator so that the differences measured reflect differences intrinsic to the protein itself, (2) the expression level of the complementing human clone can be modulated so as to restore growth but remain limiting, and (3) defined growth medium allows for the titration of exogenous cofactor (riboflavin, thiamine or niacin in this case).

At the outset we focused all efforts on assay development and variant testing for the 3 essential-gene yeast cell-based assays (human enzymes *RFK*, *FLAD1* and *TPK1*) because these presented unique challenges. Our strategy was to construct yeast strains conditional for growthy because they carried temperature-sensitive (*ts*) mutations in the essential yeast orthologs for these 3 enzyme assays. These temperature sensitive strains grow at 23-25o C, but are inviable at 37o C, thus creating conditions to assess the functional complementation of the human enzyme or enzyme variants. In addition, for the riboflavin-dependent enzymes *RFK* and *FLAD1*, we have created additional *ts* strains defective in riboflavin synthesis (deletion of the *RIB7* gene), further allowing assessment of human enzyme variants as a function of riboflavin supplementation. For all strains we determined the appropriate expression levels of human orthologs for complementation and all conditions necessary to screen enzyme variants.

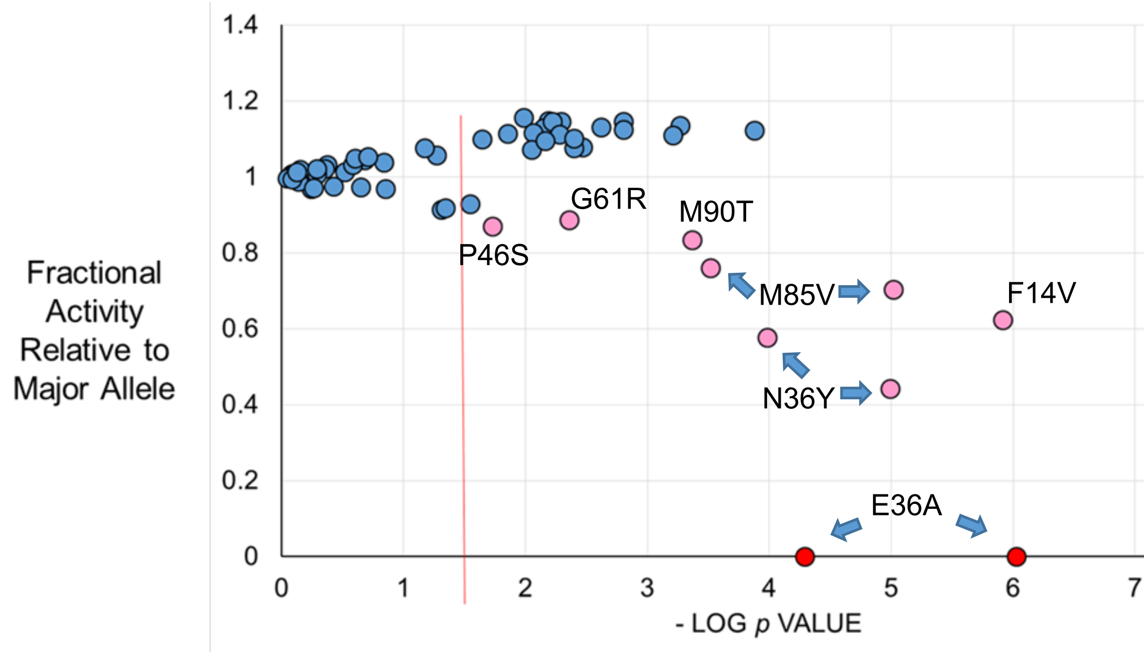
For example, the graph below demonstrates the ability of human *RFK* to complement the fmn1ts defect at 37o C as a function of riboflavin supplementation. Note how the enzyme activity is strengthened with increasing riboflavin status.



We had similarly-validated yeast cell-based complementation assay for human *FLAD1* (FAD synthase) based on complementation of the fad1ts yeast defect and for human *TPK1* (Thiamine pyrophospho-kinase) in the thi80ts background.

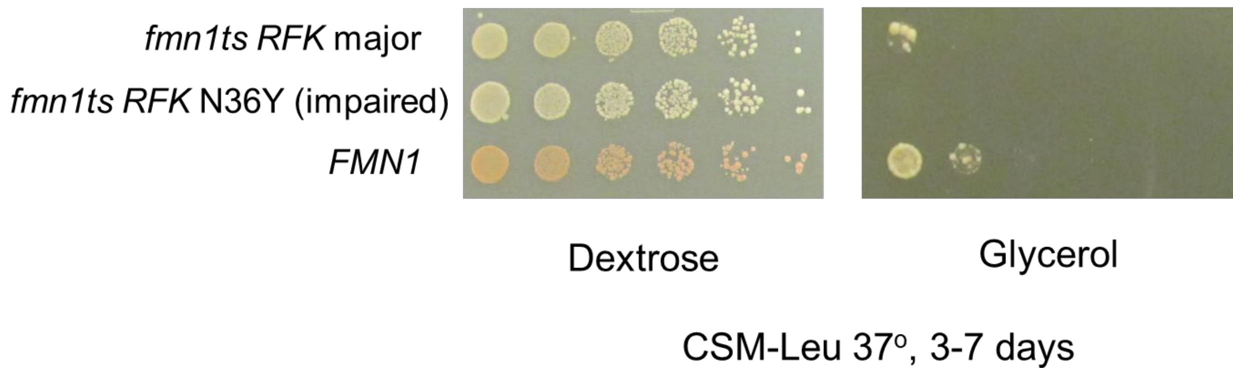
Of these 3 assays/targets, we initially focused on human *RFK*, which governs the primary step in assimilating supplemental riboflavin into the critical cofactor, flavin-adenine dinucleotide (FAD). To that end, we successfully assayed all known nonsynonymous variation in RFK as represented in the genome Aggregation Database, (or gnomAD, gnomad.broadinstitute.org) which catalogs variation from over 130,000 sequenced exomes. As depicted in the graph below, a total of 54 *RFK* variants have been assayed for functional impact and riboflavin-responsiveness. The y-axis shows the relative activity of each variant enzyme by expressing a ratio of variant activity/major allele activity (thus a value of "1" means the variant has no effect on wild-type activity). The x-axis shows the statistical confidence (as -log p-value) based on replicate measurements (vertical red line indicates p-value = 0.05).

The pink data points represent 6 naturally occurring variants (F14V, N36Y, P46S, G61R, M85V, M90T) that show measurable impact on activity but are augmented to near wild-type levels by riboflavin supplementation. Thus, >10% of all variants in *RFK* have measurable, yet remedial, impact. The red data points are duplicates of the complete loss-of-function E86A variant, which cannot be resuscitated.



Interestingly, most *RFK* variants result in a slight increase in activity over the wild-type enzyme (the majority of data points show a relative activity > 1), especially at lower p-values (-log p-value > 1.3). The significance of the mild effects of most variants on enzyme activity, as well as the tendency for slight gain-of-function will be further explored in future assays.

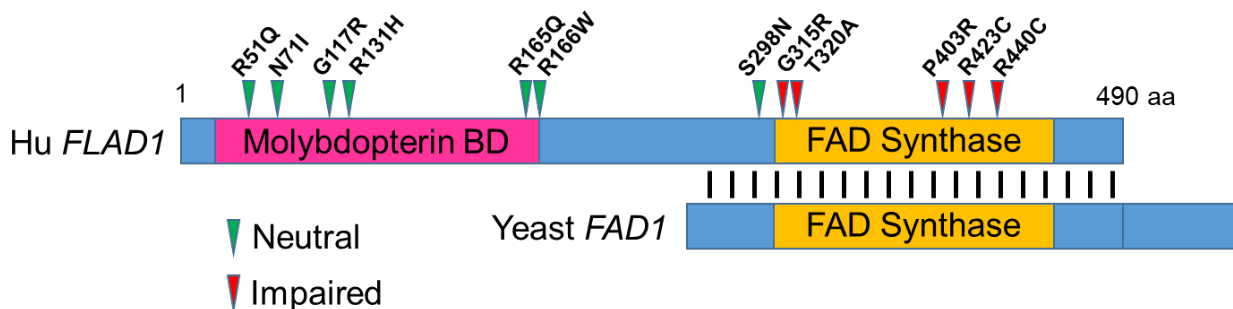
This was surprising given our previous experience with functionally annotating nearly 30 vitamin-dependent enzymes but may be significant in demonstrating that enzymes involved in cofactor synthesis (here FAD synthesis from riboflavin) may be less tolerant to functional changes than more prevalent cofactor utilizing enzymes (e.g. – *MTHFR*/folate). We will continue experimentation to validate these findings. For example, we developed a general cell-based assay for mitochondrial function by demanding that yeast cells driven human *RFK* variants grow in the presence of a non-fermentable carbon source (glycerol/ethanol). Under these conditions, growth is dependent on glycerol metabolism in the mitochondria by flavin-dependent reactions to generate ATP. The figure below shows cell dilutions on agar plates for 3 different yeast strains: wild-type (3rd line, *FMN1*), and 2 strains driven by human *RFK* (major allele, 1st line) or the impaired human variant (N36Y, 2nd line). Clearly the variant is compromised in mitochondrial function as demonstrated by its inability to grow on glycerol as carbon source. Thus, this general assay can be used to further characterize *RFK* variants for impacts on mitochondrial function and ATP generation and, importantly, whether mitochondrial function and energy stores are riboflavin-responsive.



In addition to *RFK*, we further examined variation in *FLAD1* (FAD synthase), the second step in assimilating supplemental riboflavin into the cofactor FAD. As reported last year, we tested 7 nonsynonymous *FLAD1* variants (R51Q, N71I, G117R, R131H, R165Q, R166W, and S298N), all of which were benign in our assay. We subsequently noted that these variants all occur within the N-terminal half of the enzyme, within the molybdopterin-binding domain, but outside of the catalytic core domain. We determined that frameshifts mutations in the N-terminal half impact function, which led us to hypothesize that regions outside the catalytic domain might be insensitive to point mutation. This is corroborated by clinical observations, which document clinically relevant amino acid substitutions in the catalytic domain, but only truncation/frameshift mutations in other domains of the protein.

To test the hypothesis of domain-specific functional impact, which will aid in diagnosing the sensitivity of novel variants, we assayed a second set of population variants restricted to the C-terminal catalytic domain (see figure below). We chose 5 variants predicted to be deleterious based on evolutionary conservation (G315R, T320A, P403R, R423C, R440C; see figure below), and all 5 were impaired in the yeast complementation assay. Thus, this may indeed indicate that certain domains of proteins can be more sensitive for compromising mutations (while, conversely, some domains may be tolerant to mutation).

The figure below depicts a similarity alignment between human *FLAD1* and the yeast ortholog *FAD1,* with all point mutations tested to date. The N-terminal portion of the human enzyme contains a regulatory molybdopterin binding domain that is not present in fungal enzymes and whose function is currently unknown.



<u>Expansion of Scope in Final Years of Project</u>

This project was originally designed to understand the utility of coupling cofactor supplementation to personal genetic information in a way that could have the greatest impact on peak performance and training. Consequently, this study was to focus on the potential for nutritional tuning of metabolic steps

essential for generating cellular energy stores (ATP production and mitochondrial function), which have a significant influence on physical and cognitive performance. As detailed in the Major Goals section, we have chosen to focus on 5 key enzymes involved in these processes that may be amenable to cofactor augmentation and for which common genetic variants exist that may affect enzyme function: 1) *RFK* (Riboflavin kinase), 2) *FLAD1* (FAD synthase), 3) *TPK1* (Thiamine phosphokinase). 4) *NAPRT* (Nicotinate phosphoribosyl transferase), and 5) *KYNU* (Kynureninase). For each target enzyme, we intended to interrogate all described nonsynonymous SNPs for functional impairment, metabolic impact and vitamin-remediability in functional assays. We wished to determine which changes are benign and which have a functional consequence. Further, we wished to determine which may respond to nutritional tuning via cofactor remediation. These studies were to build on our previous datasets to further understand the characteristics of remedial alleles and to develop better predictive factors from genome variation data.

However, during the course of these studies, we re-examined the widespread utility of the datasets we originally proposed to generate to inform a diagnostic test which could be applied to all members of the armed forces. While there are numerous mutations in these (and other) metabolically relevant genes which are significant in the aggregate, most are rare in the population and, thus, information on any previously identified mutation is only of limited use to those infrequent carriers.

Newer sequencing technologies have enabled far more rapid production of human exomes/genomes than imagined when this study began. Consequently, we became aware that there is an extraordinary degree of genetic variation harbored in each individual from rare or family-restricted mutations, as well as the contribution from de novo point mutations (estimated at 60/person/generation). Thus, it is inescapable that essentially all of the non-lethal genetic variation space will be seen somewhere in the population. Indeed, public databases are no longer able to stay current to annotate genetic diversity rendering it likely that the 1.4 million members of the US Armed Forces will contain many variants that have never been described, let alone tested for function. It is for this reason that the originally proposed datasets (which are derived from only currently known, annotated variants in target genes) were valid, but were becoming increasingly limited in utility.

Moving forward at this point in the project, we proposed to change the paradigm. We believed that an approach designed to generate a comprehensive classification of **all** potential variation in a metabolically important gene would be a better use of the assays developed in previous years, and of much wider utility for the Armed Forces in general. Specifically, we had been developing a process to prospectively generate an experimentally-derived **genetic atlas** in which the functional consequences of all possible missense mutations on a gene's function are measured. This strategy can effectively generate a look-up table for interpreting newly observed variants as fast as they are described for any individual through genetic testing. We will still use the same, fully-validated system of surrogate genetics (yeast complementation) to screen comprehensive libraries of in vitro-generated mutations for all target genes. **These studies would result in the generation of complete mutation maps defining the effects of the full spectrum of amino acid substitutions to enable interpretation of personal genome information**. A validated process for generating genetic atlases will produce a comprehensive and perpetual resource for any gene to which it is applied and, therefore, will be of widespread utility to all members of the Armed Forces, present and future. The 5 genes emphasized in this proposal because of their critical roles in respiration and generating cellular energy stores would be high-priority targets for gene atlas generation.

New Paradigm: Quantitative Complementation

The creation of comprehensive datasets of gene variant function (gene atlases) has been achieved in a few cases for small genes or portions of genes. However, the resource burden for this precluded the scale and cost-structure necessary for practical applications. **Therefore, we have reimagined the process to**

develop new methods that reduce input costs/effort and, importantly, reduced to practice the key new steps to validate its effectiveness.

Our process (Fig. 1) consists of three components:
1) A way of creating comprehensive mutant libraries in a vector allowing direct functional tests.
2) "En masse" functional evaluation for the function of variants by their quantitative contribution to yeast fitness; and
3) Deconvolution of data from pools of variants to assign individual scores for each variant.

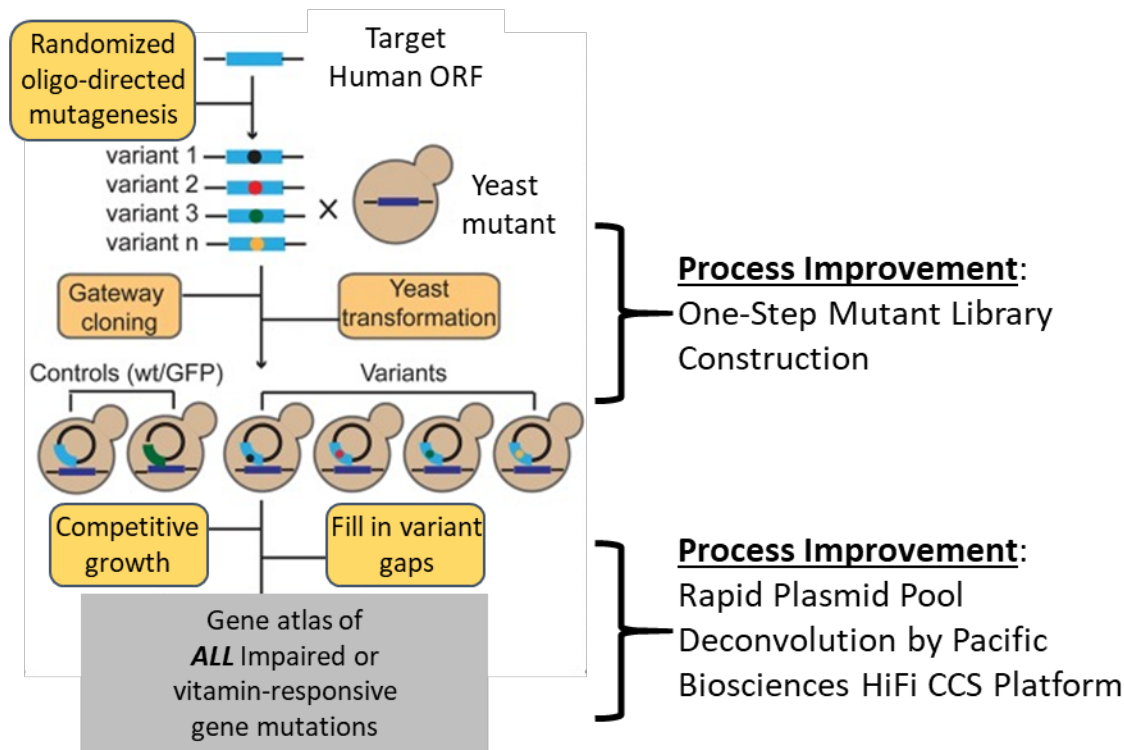


Fig. 1. *Process diagram for gene atlas construction shown for any target human metabolic enzyme. The process is applicable (and essentially identical) for many genes by simply using a different parental yeast strain allowing the appropriate complementation. Highlighted on the right (brackets) are the two technological improvements developed during the past year.*

Process Development I: Combinatorial Mutagenesis for Creating Comprehensive Mutation Libraries

As part of technology development during the final years of this project, we developed a streamlined process for a critical technology component: the generation of comprehensive mutation libraries for functional testing of enzyme variants. The process (Fig. 2) was based on the high-speed proofreading Phusion High-Fidelity DNA polymerase extension of a single phosphorylated degenerate primer annealed to a plasmid template. The extended DNA is concomitantly ligated by thermostable *Taq* DNA ligase, which enables thermocycling to create single-strand circles containing a collection of mutations after each cycle. The reaction containing single-strand DNA can directly be used to transform *Escherichia coli* after removal of the template plasmid by DpnI digestion.

The advantageous hallmarks of this process were: 1) Phusion High-Fidelity DNA polymerase that is a proofreading polymerase with extremely high extension rates, thus making the method suitable for very large plasmids, 2) the use of a thermostable DNA ligase to generate covalently closed, single-strand plasmids each cycle, and 3) the use of long mutagenic oligonucleotides that allow pooling (to target

multiple adjacent codons in one reaction) while still ensuring only one oligonucleotide binds to the plasmid template.

We validated our final protocol on multiple target codons in a prototypical vitamin-dependent enzyme: Cystathionine-beta-synthase (*CBS*). On all target loci tested, we have seen >70% success rate in generating a range of mutant codons on first pass. One such example is shown in Fig. 2. Here, the glutamine at position 400 (E400) was targeted for randomization by using a mutagenic oligonucleotide with 31 different nucleotide variations of the parent codon. Typical of our success rate, we observed 24 of the expected 31 changes after one round of thermocycling.

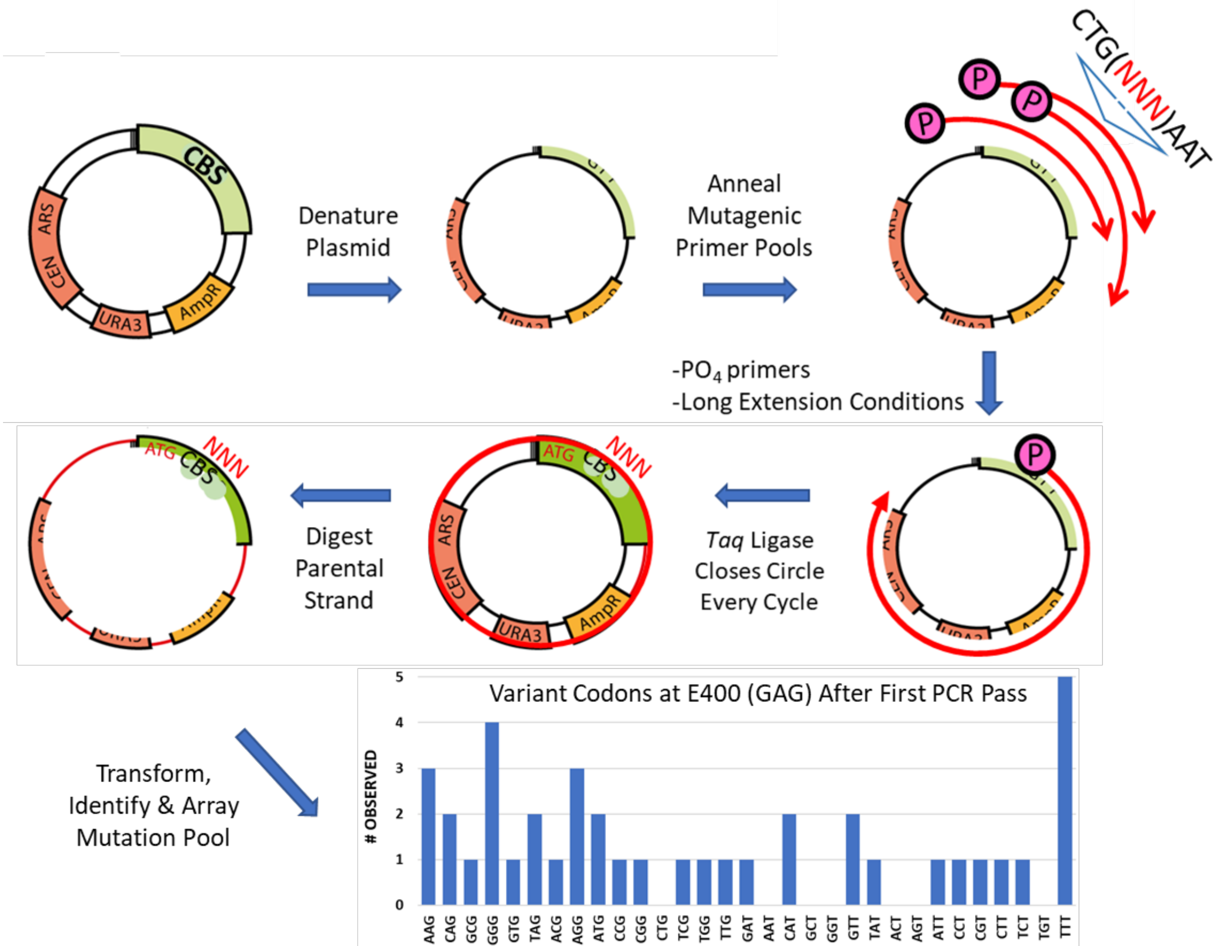


Fig. 2. *Comprehensive mutation libraries were created in a single thermocycle reaction (clock-wise from top left). Overlapping, degenerate oligonucleotides that each target a specific codon are pooled so that only one primer binds/template molecule. Conditions that ensure long, high-fidelity extensions resulted in covalently closed circles by Taq ligase. Bottom right bar graph shows example results for a target codon (Glutamine (E) at position 400) in CBS.*

Process Development II: The Transformational Impact of Long-Read Sequencing to Deconvolute Functional Pools

As summarized above, functional testing of enzyme variants is performed in a massively parallel fashion by competitive growth. In this way, many variants are grown together in a mixed pool in which functionally impaired alleles are selected against and will display lower relative abundance (decreased

fitness) over time. The end result of such fitness testing will be a complex collection of variants with different relative abundances based on function. While straightforward, this presents an analytical need to identify all pool members and, importantly, accurately assess abundance.

Much of the labor involved in providing quality control measurements during mutant library construction and subsequent fitness evaluation was caused by the three competing mismatches: 1) the need to sequence the *entire* open reading frame for each mutant constructed to be sure there was only a single mutation (not easily be accomplished by widely-used Illumina (short-read) platforms), 2) the need to easily survey complex variant mixtures, and 3) the need for low error rate. This has previously been accomplished by full-length ORF sequencing during library construction, followed by an elaborate method to uniquely tag each variant with a short barcode DNA sequence (which is ultimately what is monitored in the competitive growth assay). However, this process demands a cumbersome workload during library construction to array libraries and ensure each variant is uniquely and accurately coded. This is further compounded by QC failures inherent in the process.

To circumvent these burdens and increase efficiency, we envisioned a process in which the mutation itself is directly monitored in a method generating accurate, long DNA sequence reads (>50 Kb) to cover the entire open reading frame in a single read. To accomplish this, we adopted and formatted the recently-introduced HiFi CCS (Circular Consensus Sequence) platform from Pacific Biosciences (Palo Alto, CA), which is illustrated in Fig. 3. A simple plasmid prep from a competitively grown culture of variants is subject to HiFi CCS in which highly processive, rolling circle-based sequencing covers both strands of the plasmid multiple times ***in a single read***, thus creating a consensus sequence which greatly minimizes error. As shown in a representative experiment of 37 pooled variants of CBS (bottom left panel, Fig. 5), **this process can accurately and easily deconvolute plasmid mixtures according to abundance/function**.

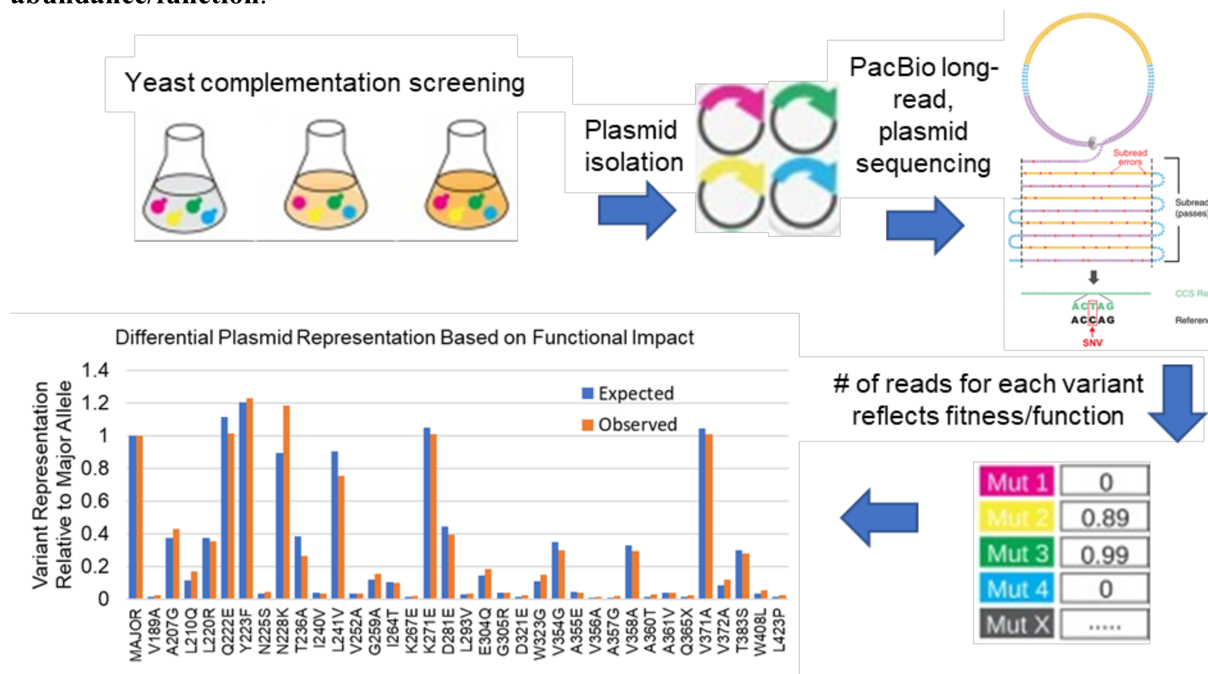


Fig. 3. *Analysis of competitive growth effects on CBS enzyme variants by a single-step plasmid isolation followed by HiFi Consensus Circular Sequence analysis (Pacific Biosciences). This process has the benefit of capturing the complete plasmid sequence in a single read, enabling removal of multiply-mutated targets from analysis, and accurately measuring relative fitness of variants by read depth. The bar graph in the lower left measured the fitness of 37 pooled CBS mutant plasmids with differential*

function. The expected relative abundance for each plasmid in the pool is shown in blue bars and relative abundance observed by PacBio read counts in orange bars.

There were several advantageous hallmarks for this process improvement. First, the process is relatively manipulation-free (and, therefore, highly efficient): only a plasmid preparation is necessary following competitive growth. Second, there are no amplification steps, thereby eliminating PCR-based errors or biases. Third, sequence capacity in a single run will allow for the simultaneous analysis of 1,000+ variants. Lastly, by capturing the entire ORF in a single read, multiply-mutated plasmids can be informatically filtered, thus focusing only on the desired single-nucleotide changes.

Thus, these 2 process improvements, which we have made significant progress on during the last phase of this grant, synergize in a highly efficient pipeline necessary for the scale-up and broader applicability to additional vitamin-sensitive gene targets including the ones in this study.

Conclusions/Wrap-Up Plans

Two developments have fundamentally changed the landscape regarding the delivery of personal genetic information: The enormous success of consumer-driven endeavors such as Ancestry and 23andMe, and the increasing adoption of genome sequencing in clinical care. To date, however, the tools available for interpreting genetic information are still focused on disease-causing variation, ignoring the millions of variants of unknown significance each individual harbors.

Our goal moving forward is to produce an interpretive tool, the gene atlas, which defines **all** functionally significant variation in relevant genes and serves as a look-up tool to inform individual genetic profiles. Beyond clinical support for members of the armed forces, we will enable unprecedented insight into genetic variation and metabolic impact, along with the potential for nutritional remediation to enhance performance.

Whereas the original proposal described the interrogation of known variation within a set of genes critical for mitochondrial and respiratory function, our recent technology advance will provide functional impact data for all possible single nucleotide changes in these and other high-value target genes/enzymes.

This study provided several valuable findings and developments in the continued effort to bring value to genetic testing as a means for the Armed Forces to enhance individual performance safely and effectively. First, we determined the functional impact and cofactor-responsiveness for all known variation in the FAD biosynthetic pathway providing additional content for a genetic screening program. Second, we further validated the yeast complementation assay as a valuable, relevant and versatile tool for studying the impact of genetic variation in critical metabolic pathways. Lastly, and arguably most importantly, these studies provided a springboard to investigate more automated, higher-throughput versions of comprehensive scanning mutagenesis coupled with variant functionation (such as the automated CRISPR-based editing platform provided by Inscripta (Boulder, CO).

With further validation, this will be applied to the target enzymes described in this original proposal that are critical to mitochondrial function and maintaining cellular energy stores, specifically riboflavin and thiamine-mediated steps. RFK (riboflavin kinase) and FLAD1 (FAD synthase) are attractive targets because their function is easily measured in yeast complementation assays.