Humanization of the sterol biosynthesis pathway in yeast

Michelle Vandeloo

A

Thesis

in the

Department of Biology

Presented in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

at

Concordia University

Montréal, Québec, Canada

December 2022

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: Michelle Vandeloo

Entitled: Humanization of the sterol biosynthesis pathway

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Michael Sacher

_____ Examiner
Dr. Vincent Martin

_____ Examiner
Dr. Michael Sacher

_____ External Examiner
Dr. Malcolm Whiteway

_____ Supervisor
Dr. Aashiq Kachroo

Approved by

_____ Graduate Program Director
Dr. Robert Weladji

_____ Dean of Faculty
Dr. Pascale Sicotte

2022/2023

Abstract

Humanization of the sterol biosynthesis pathway in yeast

Michelle Vandeloo

Baker's yeast, or *Saccharomyces cerevisiae*, shares thousands of genes with humans, making it invaluable for studying human gene function, genetic interactions, and evolution of critical cellular processes. Humanized yeast, i.e., yeast with human genes functionally replacing their yeast orthologs, enables the direct study of human gene function and their disease-causing variants in a simplified model. Recent systematic functional complementation assays performed one gene at a time revealed that genes belonging to pathways or complexes are either entirely replaceable or are not. In the case of genes that are not easily replaceable, we hypothesize that some essential interactions have evolved in a species-specific manner, resulting in incompatibility. Therefore, humanizing whole processes may reveal incompatibilities towards humanization, by restoring local genetic or physical interactions, allowing replaceability of those genes. We tested this approach to humanizing the sterol biosynthesis pathway in its entirety in yeast. Using marker-less CRISPR-Cas9 selection and Homology Directed Repair, we show that several human genes can replace their yeast equivalents at their native loci generating single-gene humanized strains. Next, we demonstrate that multiple yeast genes belonging to the sterol biosynthesis pathway are replaceable in a single strain using a sequential approach (10 of 16 genes). Characterization of the humanized sterol strains reveals the impact of humanization on the fitness of yeast such as altered growth rates and temperature sensitivity, and proteome-scale changes affecting other biosynthesis pathways. Finally, we design a novel strategy to increase the efficiency of functional replaceability to provide a clear readout of replaceability. In future, these strategies will enable building an entirely human sterol biosynthesis pathway in yeast.

Acknowledgements

I would like to first thank my supervisor, Dr. Aashiq Kachroo, for supporting this work, and for his guidance throughout the project.

Next, I would like to thank my colleagues in the Kachroo lab, for welcoming me to the group during the stressful time of lockdowns, and helping me along the way with thoughtful discussions, troubleshooting tips, and teaching me new skills as I explored the field of yeast genetics and synthetic biology. Thanks to: Mudabir Abdullah, Brittany Greco, Devina Singh, Farhat Zafar, Dr. Jeff Bouffard, Aaliya Naaz, Chris Dykstra, Joseph Trani, and Homin Jeong.

On a technical level, I would like to thank Dr. Eric Bonneil at l'Université de Montréal for his help with the LC/MS; Dr. Chris Law at the Centre for Microscopy and Cellular Imaging at Concordia for training me fluorescence microscopy; and Dr. Mike Hallett for teaching me how to code and utilise bioinformatics packages and pipelines, a skill that came in handy more than I expected it to in writing this thesis.

I would also like to thank Dr. Vincent Martin and Dr. Michael Sacher for being on my committee, taking time out of their busy days to provide me with thoughtful feedback and ideas over the course of the project.

A special thanks to Michelle Abramowitz, Lily Hullinger, and Farhat Zafar for proof-reading my thesis, and providing encouraging comments in the margins throughout.

Finally, and perhaps most importantly, I would like to thank my friends for patiently smiling and nodding as I got excited about incredibly technical details of my project; and my family, for their unending support and encouragement for me to keep taking interesting opportunities and doing interesting things as I develop my career in science.

Contribution of Authors

This project was conceived by Dr. Aashiq Kachroo. Several humanized sterol biosynthesis strains were engineered by Dr. Aashiq Kachroo, Sophie Curie, and Riddhiman Garge at the University of Texas at Austin. CRISPR plasmids in the ScEDIT collection were built by Mudabir Abdullah, Brittany Greco, Dr. Aashiq Kachroo, Leanne Bourgeois, and Michelle Vandeloo. Engineered strains tested by ScEDIT plasmids were built by Mudabir Abdullah, Brittany Greco, Sophie Curie, and Michelle Vandeloo. Sample preparation of the humanized alpha-proteasome strains for WGS was done by Mudabir Abdullah. All other experiments were conducted and analysed by Michelle Vandeloo.

Table of Contents

List of Figures vii
List of Tables viii
Chapter 1 – Background 1
1.1 Yeast as a model organism for the study of human biology 1
1.2 Scaling humanized genes up to humanized pathways in yeast 3
1.3 The sterol biosynthesis pathways in yeast and humans 9
1.4 CRISPR-Cas9 gene editing in *S. cerevisiae* 15
1.5 Study goals 17
Chapter 2 – Materials and Methods 20
2.1 Gene-phenotype association data curation and code 20
2.2 Strains and media 20
2.3 Plasmid construction 21
2.4 *E. coli* transformation 22
2.5 CRISPR-Cas9-mediated humanization at the native yeast locus 22
2.6 PCR genotyping of engineered loci 24
2.7 Sequencing confirmation of humanized loci 25
2.8 Whole genome sequence analysis parameters 25
2.9 Growth assays 26
2.10 LC/MS sample preparation and data analysis 27
2.11 Microscopy 28
2.12 Colony forming unit assays 28
Chapter 3 – Humanizing the sterol biosynthesis pathway in yeast 29
3.1 Design and testing of CRISPR-Cas9 plasmids 29
3.2 Exogenous PCR repair template for humanization 32
3.3 The ScEDIT toolkit for CRISPR-Cas9-based gene editing in *S. cerevisiae* 36
3.4 Engineering pathways in yeast 39
3.5 Partially humanized sterol pathways enable the replacement of individually non-replaceable genes at the native locus 42
3.6 A novel, higher-efficiency strategy for gene humanization 43
Chapter 4 – Characterization of humanized strains to identify pathway engineering bottlenecks 48
4.1 Utility of characterizing strains 48
4.2 Growth behaviour of humanized sterol strains 49
4.3 Differential protein expression in humanized strains 57
4.4 Differential localization of human and yeast orthologs expressed in yeast 65
Chapter 5 – Conclusions and Future Directions 70
5.1 Future directions 70
5.2 Conclusion 74
Chapter 6 – Bibliography 76
Supplementary Information 83

List of Figures

Figure 1: Humanized yeast is an advantageous system to model disease. 5
Figure 2: Genetic modules govern functional replaceability. 6
Figure 3: Overview of ergosterol biosynthesis and key intra-pathway regulators. 12
Figure 4: Sterol biosynthesis genes in yeast are functionally replaceable by their human orthologs. 18
Figure 5: CRISPR-Cas9-based humanization approach in yeast. 30
Figure 6: CRISPR-Cas9 plasmids induce lethal DSBs in essential yeast genes. 32
Figure 7: Novel individually humanized genes at the native locus in the sterol biosynthesis pathway in yeast. 35
Figure 8: The ScEDIT collection of CRISPR-Cas9 plasmids for yeast genome editing. 38
Figure 9: Sequential humanization of multiple genes in the sterol biosynthesis pathway in a single yeast strain. 41
Figure 10: Ectopic chromosomal repair template increases HDR efficiency over exogenous PCR repair template for replaceable genes. 46
Figure 11: Growth behaviour of single-gene humanized strains. 50
Figure 12: Growth behaviour of multi-gene humanized strains. 53
Figure 13: Illustrative simulation of the multiplicative and cooperative models for higher order humanization. 55
Figure 14: Spotting assays of humanized strains. 57
Figure 15: Detection of native yeast and replaced human proteins in humanized and wild-type strains. 59
Figure 16: Differential expression of yeast and human sterol biosynthesis proteins in humanized strains. 60
Figure 17: Altered individual protein expression in humanized strains. 62
Figure 18: KEGG pathway enrichment in altered protein expression in humanized strains. 64
Figure 19: Differential localization of human and yeast orthologous proteins in *S. cerevisiae*. 69
Figure 20: Humanization of the sterol biosynthesis pathway reveals bottlenecks towards whole-pathway engineering 73

Figure S1: Whole genome sequencing reveals precise replacement of multiple genes in a single strain 95
Figure S2: Clustered heatmap of differentially expressed proteins in humanized yeast. 96

List of Tables

Table 1: Mendelian disease and phenotype associations with genes of the upstream cholesterol biosynthesis pathway. 13
Table 2: CRISPR sgRNA target sequences expressed on Cas9 plasmids. 31
Table 3: Humanized strains with a single sterol biosynthesis gene replaced to date. 34
Table 4: Humanized strains with multiple sterol biosynthesis genes generated to date. 40
Table 5: Growth parameters of humanized sterol biosynthesis strains 54
Table 6: Localization of GFP-tagged yeast and human proteins in their respective cell types. ... 65

Table S1: Primers to confirm presence and absence of human and yeast genes by PCR and sequencing 83
Table S2: Primers for amplifying repair template with 80bp homology arms 86
Table S3: Human gene coding sequence templates 89
Table S4: Unique SNPs from WGS of 5-gene humanised alpha proteasome strain 90
Table S5: ScEDIT collection sgRNA sequences cloned into CRISPR plasmids 91
Table S6: Golden Gate primers for assembly of repair template in LEU2 plasmid YTK97. 93
Table S7: Amino acid sequence identity alignment between yeast and human orthologs of the sterol biosynthesis pathway. 94

Chapter 1 – Background

1.1 Yeast as a model organism for the study of human biology

For decades, budding yeast (*Saccharomyces cerevisiae*) has served as a versatile and pioneering model organism for studying eukaryotic cell biology and genetics[1]. For brevity, *yeast* and *S. cerevisiae* are used interchangeably throughout the thesis. Many innate qualities of yeast make it ideal for experimental systems: It has a fully-sequenced and well-annotated genome of approximately 6000 genes, it grows easily and rapidly in rich media, it exists in both haploid and diploid states, and it is highly tolerant of genetic manipulation[1–3]. In addition to its ability to maintain and read genes from plasmids and artificial chromosomes, yeast also has uniquely high rates of homologous recombination compared to non-homologous end-joining (NHEJ)[4]. This makes the introduction of targeted mutations and integration of heterologous DNA sequences into the genome highly effective, and contribute to yeast being an attractive organism for genetic engineering with industrial and research applications[5]. These qualities have enabled the development of a wealth of toolkits and collections facilitating the study of biological processes in yeast, and extensive database collections to host these annotations for *in silico* study of yeast[6,7]. These collections include applications for conditional gene deletions and knockouts[8], fluorescence-microscopy protein detection[9], controllable gene expression systems[10–12], next-generation cloning and expression vectors[13], and genetic interactions[14].

The high genetic tractability and rapid growth of yeast strains make them desirable platforms for metabolic engineering and synthesis of products such as biopharmaceuticals, biofuels, chemicals, and cannabinoids[15–18]. Plasmid vectors—either with low-copy origin of replication genes (centromere or CEN vectors) or the high-copy 2-micron origin—and an array of well-characterized and titratable yeast promoters, allow for careful control of copy number in yeast[19,20].

In particular, thanks to the efficient homologous repair machinery, heterologous gene integration at the chromosomal level is also possible, which has led to the development of tools such as the landing pad system that controls gene copy number within the genome[21]. Genetic manipulation of yeast can be scaled from single genes up to heterologous construction of metabolic engineering pathways, introducing novel challenges that require an understanding of how multiple genes interact[19,22]. While the high manipulability of yeast makes them ideal as cell factories, the large number of highly conserved genetic systems between yeast and humans also makes them ideal for the study of human biology and disease.

Despite a billion years of evolution, the yeast and human genomes still retain a high level of functional conservation, with many genes retaining similar functions[23–25]. In particular, genotype-phenotype associations and genetic interactions, difficult to study systematically in humans due to complexity and variability between systems and cell types, are far better mapped out in yeast[26–28]. While the human genome has now been entirely sequenced, it remains difficult to systematically screen phenotypes and interactions in human cells due to the heterogeneity of gene expression across cells, as well as in response to genetic factors, which makes yeast an attractive model towards mapping out conserved interactions in human cells[1,26,29]. Yeast is an effective platform upon which to study human biology and evolution, model disease phenotypes, and screen novel drug candidates[27,30–32]. Even discoveries related to highly heterogenous diseases like cancer and severe mitochondrial genetic diseases, as well as the complex effects of aging, have been elucidated thanks to the awesome power of yeast genetics[33–35].

1.2 Scaling humanized genes up to humanized pathways in yeast

Thanks to the many orthologous genes between yeast and humans, as well as the ease of genetic modification, yeast can serve as a platform to study human biology not only through study of conserved processes, but also by engineering yeast that use functionally complementing human genes. This systematic rewiring of model organisms to express orthologous human genes and cover essential yeast gene functions is referred to as "humanization". Besides yeast, humanization is being investigated in other higher eukaryotic models like flies, worms, and mice to make these model organisms more accurate platforms to study basic human biology and disease[24,36,37]. Humanization of yeast is achieved by knocking out or downregulating the yeast gene while a human gene is expressed on a plasmid, allowing direct investigation of the functional conservation of genes between species, rather than only at the sequence level[24,38]. For decades, the ability of human genes to complement their yeast orthologs has been investigated as a way to identify functionally conserved proteins separated by a billion years of evolutionary divergence[39]. However, recent systematic screens have revealed hundreds of functionally replaceable human genes in yeast[24,38,40–45].

Large scale complementation assays have fundamental implications and are powerful tools to study evolutionary divergence between yeast and human genes[38,40]. In the years since, research has unveiled more than 700 yeast genes that are functionally replaceable by their human orthologs[1]. Thanks to many systematic gene deletion, temperature-sensitive and chemogenomic assays, the number of unique genotype-phenotype relationships in yeast far surpasses those known in humans, again making yeast attractive models in which to study altered phenotypes of mutant human genes (Figure 1A). Furthermore, many replaceable human genes have variants of unknown clinical significance, as well as Mendelian and rare disease associations, that could be studied in a

rapid and high-throughput manner in humanized yeast assays (Figure 1B)[1,41]. Humanized yeast has been used to study effects of polymorphisms in genes associated with diseases such as Alzheimer's, Parkinson's, neural tube defects, and motor neuron diseases[46–49]. They are also excellent platforms for novel drug discovery, providing rapid assessment of phenotype changes in response to large-scale drug screens[43,50,51]. All of these assays limit confounding factors of gene expression and genetic interactions, but they themselves are limited in that they do not capture the larger cellular context of these human genes, particularly within pathways or complexes[1]. Therefore, there is a need for not just yeast with humanized genes, but for yeast with entirely humanized complexes and pathways to maintain human-specific local genetic and physical interactions[23].

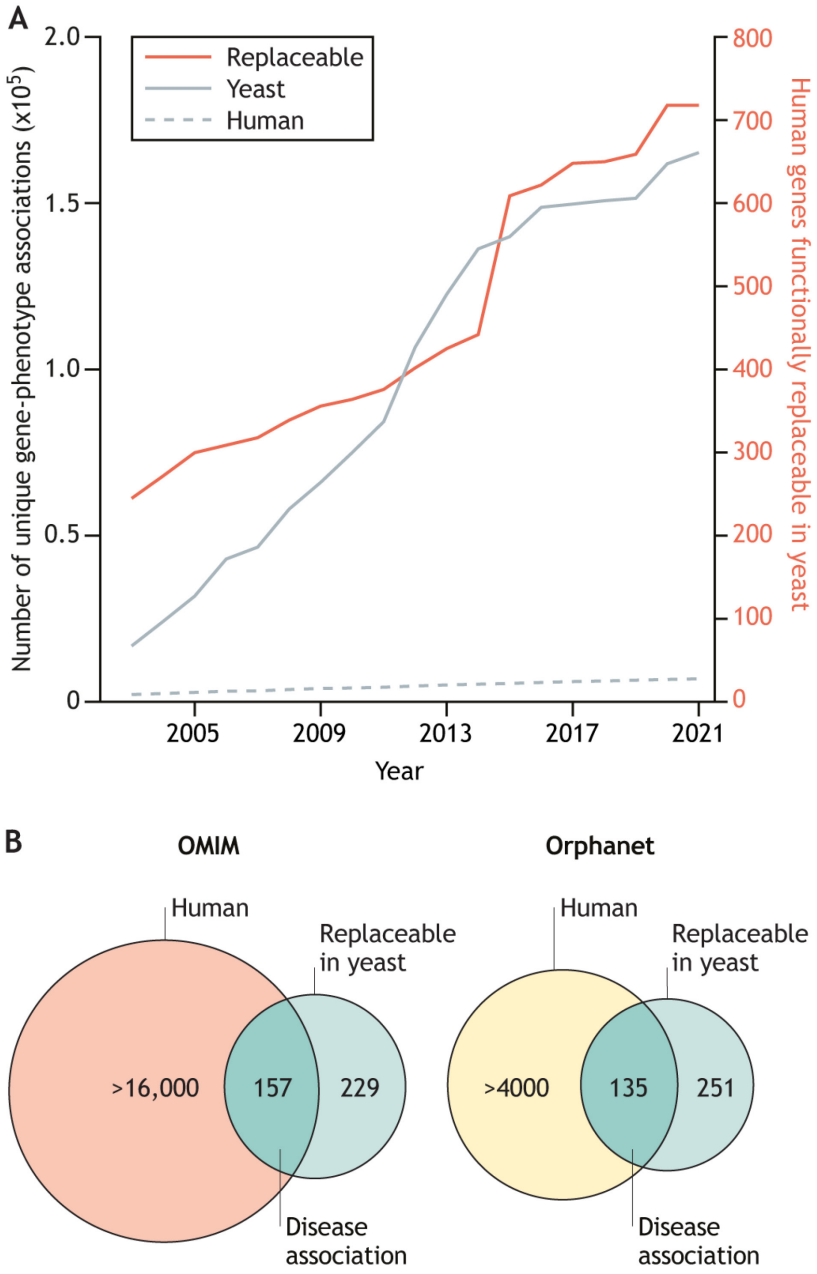


Figure 1: Humanized yeast is an advantageous system to model disease

(A) The number of known single genotype–phenotype associations in yeast (gray solid; left y-axis) has surpassed those known in humans (gray dashed; left y-axis). Similarly, many human genes are functionally replaceable in yeast (red; right y-axis). (B) Humanizable essential genes in yeast overlaid with genes from the Online Mendelian Inheritance in Man (OMIM) database. Of the 386 essential yeast genes that can be functionally replaced with human orthologs, 157 have disease association in the OMIM database. Of the 386 functionally replaceable essential yeast genes, 135 overlap with rare disease-associated genes in the Orphanet database. *From Kachroo et al 2022* [1].

Compared to humanizing individual yeast genes, humanizing entire pathways and complexes represents a far more sophisticated engineering challenge. An entirely humanized pathway requires all these genes to be replaced in a coordinated fashion, where they are expressed and functional collectively rather than just as an individual gene. Particularly when genes are replaced sequentially, there are many intermediary steps with questionable viability and fitness (Figure 2). While high-risk, the endeavor to humanize complete yeast pathways and complexes will provide platforms that better capture the functional contexts of human proteins, as well as provide novel ways to study genetic or physical interactions and regulation[23]. Given that the functional replaceability of human genes in yeast is highly modular, replaceable genes are usually found within modules comprising other replaceable genes. – Thus, the modularity paradigm suggests that certain pathways such as heme and sterol biosynthesis, and complexes such as the proteasome, are suitable candidate systems for full humanization[24,52].

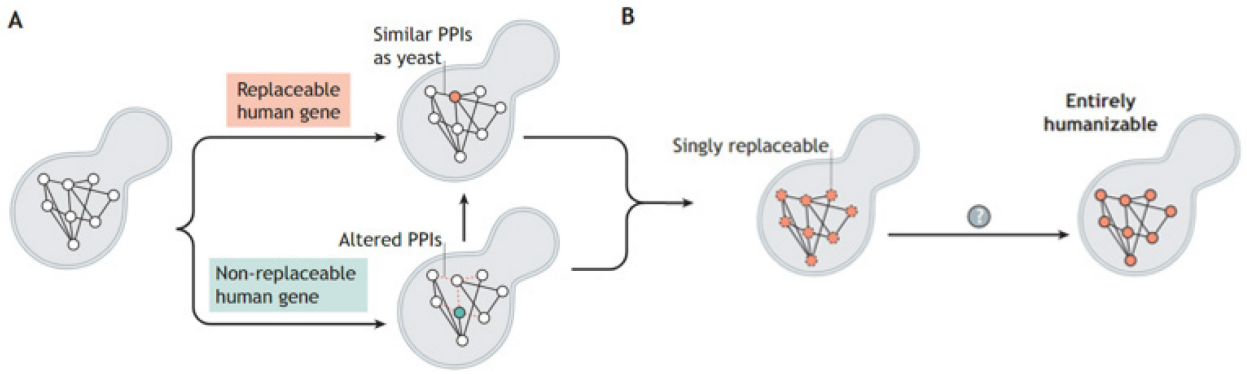


Figure 2: Genetic modules govern functional replaceability

(A) Lack of functional complementation by a shared human gene in yeast could be attributed to the inability of a human gene to perform critical genetic interactions or protein–protein interactions (PPIs) in yeast. Modifying a non-replaceable human gene to complement the yeast ortholog should allow the discovery of critical interactions or other factors, such as diverged mechanisms or regulation between humans and yeast. (B) Genetic modularity is a feature that strongly predicts replaceability and allows researchers to test whether higher-order humanizations of yeast are possible. Some modules can be humanized because most of the individual genes within the module are replaceable, either sequentially or by expressing all humanized components simultaneously. *From Kachroo et al 2022* [1] .

To date, few attempts at humanizing entire pathways or complexes have been performed, but as access to high-throughput methods and toolkits increases, the process will become more feasible for general use. One of the first successfully humanized complexes in yeast was the nucleosome, which was performed by systematically replacing yeast histones with human versions[53,54]. The humanized nucleosome yeast grew very poorly, suggesting that even small divergences in sequence are highly specific from species to species. Humanization of the adenine biosynthesis pathway in yeast enabled the discovery of novel key regulatory elements, as humanization of the Origin Recognition Complex (ORC) in yeast equally elucidated a unique insertion differentiating the selectivity of yeast and human ORC[55,56]. Recently, exploration of yeast with entirely humanized pathways and complexes has also revealed critical amino acids that enable functional replaceability of otherwise non-replaceable human genes in yeast, providing insight into key steps in the evolution of highly diverged eukaryotic lineages and species-specific functions of conserved genes[57,58]. Of particular interest is the first attempt at humanizing an essential pathway, glycolysis, in yeast, revealing evolutionarily important mutations that confer function of human genes in yeast systems[58]. Due to the immense interest in using yeast as platforms for studying human genes, attempts have likewise been made to humanize post-translational modification machinery to make functional human proteins in yeast[59–61]. While humanized pathways and complexes in yeast introduce new challenges, including severe fitness defects, the success of these attempts suggests that many of these modules are not only highly conserved at a sequence level, but at a functional and regulatory level as well. They also serve as effective platforms for the study of human biology, genetic interactions, and variant and drug screens for human genetic diseases[23,62,63].

Each of these attempts have revealed unique benefits of humanization of entire pathways, while identifying unique challenges. Most attempts at humanizing pathways in yeast, much like methods to replace individual genes, have relied on plasmid shuffling or plasmid-based complementation of multiple yeast genes with their human counterparts. While these methods are rapid and accurate, they are subject to plasmid instability, and the resulting human genes may not mimic the fine-tuned expression levels of the native yeast genes, which can be particularly important when humanizing entire systems[52]. Using yeast native promoters and terminators retains native regulation and expression of the humanized genes, which can be an advantage over the plasmid-based expression[55]. While site-specific chromosomal integration of DNA parts and recombinant genes has historically been inefficient to scale, the advent of gene editing tools such as CRISPR-Cas9 has allowed for the development of scalable and higher genome editing techniques[13,64]. Other strategies like synthetic genetic array analysis have been used to study combinations of engineered loci[14]. These methods can be multiplexed for simultaneous chromosomal engineering of multiple genes – however, they are limited by the number of selectable markers available[13,65]. GFP-tagging of multiple engineered genes and screening for increased fluorescence in strains with several engineered loci, while successful, may alter the fitness of strains due to the expression of many fluorescent proteins[65,66]. Recently, our lab has developed a novel method to humanize entire systems called MERGE (Markerless Enrichment and Recombination of Genetically Engineered loci), which harnesses the rapid mating and sporulation abilities of yeast and CRISPR-Cas9-induced gene drives to systematically combine engineered loci into a single strain[67]. This method was successfully used to humanize six of seven subunits of the proteasome alpha core in yeast.

Much like the proteasome core, the sterol biosynthesis pathway is highly conserved at a functional level between yeast and human. Previous assays using repressible promoter expression, temperature-sensitive alleles, and genetic assortment have shown that fourteen of the sixteen upstream sterol biosynthesis genes in yeast are functionally replaceable with their human orthologs[24]. While the sterol biosynthesis pathways in yeast and humans are both well characterized from a catalytic and mechanistic point of view, genetic or physical interactions are still not well understood, particularly in humans.

Using a sequential engineering strategy, my project aims to fully humanize the sterol biosynthesis pathway in yeast to provide novel ways to study the evolution of the genes in this pathway, their genetic or physical interactions, and associated diseases and drug targets.

1.3 The sterol biosynthesis pathways in yeast and humans

Sterols are essential lipid molecules synthesized by most eukaryotes, widely used throughout the cell for membrane integrity, permeability, membrane-bound enzyme regulation, and cell proliferations[68–70]. Different kingdoms have evolved to synthesize and use different sterols (i.e., ergosterol in fungi, compared to cholesterol in animals) but sterols largely retain the similar structure and function, and are expressed mainly in the cellular membranes[71]. The sterol biosynthesis pathways in yeast and humans require four acetyl-CoA molecules to produce one molecule of ergosterol or cholesterol using >20 enzymes, most of which yeast and humans have in common[72]. In humans, cholesterol is synthesized via the biosynthesis pathway and absorbed from dietary intake[72,73]. In yeast, ergosterol is only synthesized in aerobic environments as the downstream pathway requires oxygen – in anaerobic environments, yeast become auxotrophic for sterol, taking it from the environment to integrate into their membranes for essential functions[74,75].

Despite the similar functions in the cell, cholesterol is noted to be a more energetically efficient molecule, suggesting that the divergence of these pathways may have accompanied the evolution of animal and fungal kingdoms[76].

Biosynthesis of sterols in yeast and humans proceeds in two sub-pathways. The first is mevalonate and isoprenoid biosynthesis, which takes place in the vacuole and mitochondria of yeast, and between the cytosol, peroxisomes, and endoplasmic reticulum in humans[68,77,78]. This pathway is essential for the biosynthesis of isoprenoids, which serve a diverse array of functions in cells and are precursors of sterols, heme, quinones, prenylated proteins, and numerous hormones in humans[78–80]. Broadly, two acetyl-CoA molecules are condensed and transformed into mevalonate, which is further transformed into farnesyl pyrophosphate (FPP), an important intermediate for many metabolites and post-translational modification signal[81–83]. The second part of the pathway is exclusively dedicated to sterol synthesis, and begins with the condensation of two FPP molecules into squalene, a highly hydrophobic isoprenoid molecule. Squalene is further cyclized into lanosterol and zymosterol – this is the last common metabolite of yeast and human sterol pathways, and is also the first sterol that can be localized to the cellular membrane[73,75]. The final steps of the pathway include five genes that have diverged to produce ergosterol and cholesterol for yeast and humans, respectively, which are then transported to required cellular membranes[70,71,84]. The later steps of the pathways largely take place in the endoplasmic reticulum membranes, due to the high hydrophobicity of the substrates[68,71,72].

Sterol biosynthesis is a highly regulated process in both yeast and humans, regulated by intracellular metabolites and by other proteins. While sterols are essential molecules for life, they can cause problems if they or their intermediates accumulate to too high a degree (Figure 3)[85,86]. The pathway is regulated by product feedback inhibition, where high levels of sterols in the cells

will downregulate the genes for sterol biosynthesis, and low sterol content in cells triggers an activation in sterol biosynthesis [87]. Acetoacetyl-CoA thiolase is one such enzyme, regulated at a transcriptional level by intracellular sterol levels[88].The rate-limiting step is HMG-CoA (3-hydroxy-3-methyl-glytaryl-coenzyme A) reductase (*HMG1/HMG2* are paralogs with the same catalytic function in yeast; *HMGCR* is the sole ortholog in humans), which is one of the most highly regulated genes in the cell [89]. This enzyme is regulated by both intracellular sterol levels and by FPP[85]. In yeast and humans, squalene monooxygenase (*ScERG1/HsSQLE* – henceforth the prefix "Sc" denotes a *S. cerevisiae* gene and "Hs" denotes a *H. sapiens* gene) represents a further rate-limiting step, and along with HMG-CoA reductase, is regulated by similar transcription factors[72,85]. However, despite the importance of cholesterol biosynthesis in human health and disease, much remains to be learned about the regulation of this pathway, particularly involving genetic interactions among pathway members in humans[90,91]. Yeast remains a powerful tool for directly studying human genes, but only by humanizing multiple genes and entire pathways can we recreate the human-like local interaction network in the cholesterol synthesis pathway.

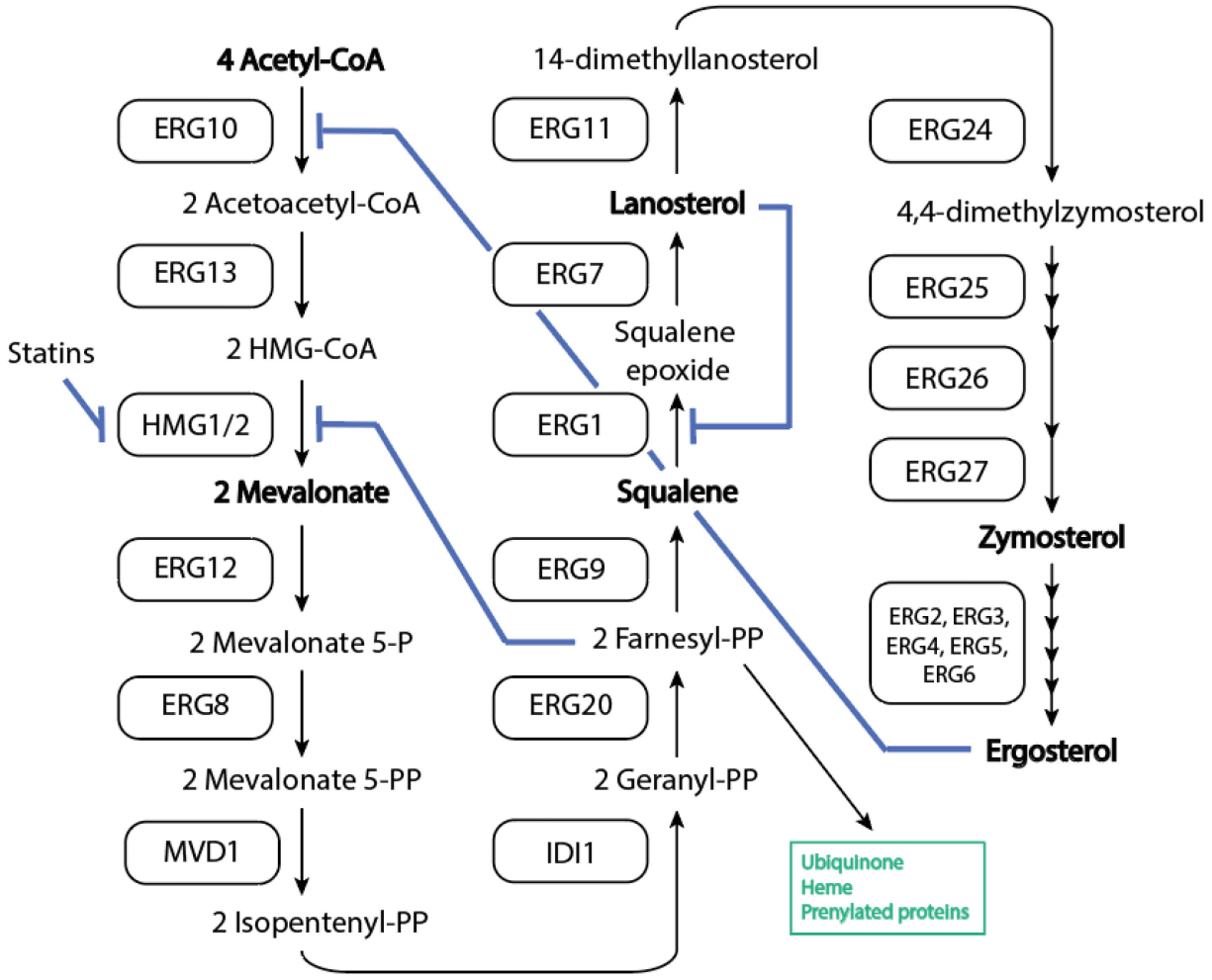


Figure 3: Overview of ergosterol biosynthesis and key intra-pathway regulators

Ergosterol biosynthesis in *Saccharomyces cerevisiae* is a multi-step pathway, beginning with Acetyl-CoA molecules that are transformed into zymosterol over >20 reactions catalyzed by 16 enzymes. The five remaining steps convert zymosterol to ergosterol. Black arrows indicate enzyme-catalyzed reactions; boxes contain enzyme names. Blue lines represent regulatory inhibition of enzymes, either by feedback inhibition within the pathway, or by external molecules. The upstream pathway is shared between ergosterol and isoprenoid biosynthesis – farnesyl-PP molecules can be further converted into sterols, or taken into other pathways (green box). *Adapted from Jordá and Puig*[85].

Mutations or regulatory disruption of sterol biosynthesis has significant consequences on organisms, resulting in a number of diseases in humans, and lower resistance to stress in yeast[86,92–94]. Loss of function mutations in most sterol biosynthesis genes in yeast are lethal, particularly those in the early mevalonate/isoprenoid biosynthesis pathway, with the exception of *Δerg24*, which is viable only when grown on synthetic complete medium[95]. The five downstream genes

that are responsible for the transformation of zymosterol into ergosterol (*ERG2* through *ERG6*) are conditionally non-essential in aerobic conditions, but mutations in these genes can lead to altered stress tolerance and sterol composition in yeast cell membranes[95]. Due to the essentiality of this biosynthesis pathway, several genes in this pathway are drug targets for effective management of fungal pathogens (i.e. *ERG11,* by azole drugs, as well as *ERG4, ERG5* and *ERG6* for novel drugs as these enzymes have no human equivalents)[96].

In humans, the regulation of sterol biosynthesis genes is essential to good health, and mutations in sterol genes are associated with disease phenotypes or known genetic diseases[93,97]. In addition to numerous genetic metabolic and autoimmune disorders (Table 1), dysregulation of cholesterol biosynthesis or metabolism is associated with many types of cancers and cardiovascular diseases[98]. Yeast models, and particularly humanized yeast, can be harnessed to characterize variants of unknown significance, as well as mutations in these rare diseases by measuring simple phenotypes such as growth rates in precisely mutated human genes[1,62]. While models of most simple genetic diseases in this pathway are possible, a model to study porokeratosis, a disease which has been linked to mutations in multiple sterol biosynthesis genes, sometimes simultaneously, would require the ability to study the genetic variance of all of these genes at once[99]. For this approach, a fully humanized sterol biosynthesis pathway in yeast could represent a novel and ground-breaking model in which to study complex genetic diseases.

Table 1: Mendelian disease and phenotype associations with genes of the upstream cholesterol biosynthesis pathway

Disease/phenotypes	Implicated gene(s)	Yeast ortholog	OMIM number	Reference
Alpha-methylacetoacetic aciduria	*ACAT1*	*ERG10*	203750	[100]
Hyper-IgD syndrome	*MVK*	*ERG12*	260920	[101]

Disease/phenotypes	Implicated gene(s)	Yeast ortholog	OMIM number	Reference
Mevalonic aciduria	*MVK*	*ERG12*	610377	[102]
Porokeratosis	*MVK, PMVK, MVD, FDPS*	*ERG12, ERG8, MVD1, ERG20*	175900	[99]
Squalene synthase deficiency	*FDFT1*	*ERG9*	618156	[103]
Alopecia and hypertrichosis	*LSS*	*ERG7*	618275; 618840	[104]
Pelger-Huet anomaly	*LBR*	*ERG24*	169400	[105]
Greenberg skeletal dysplasia	*LBR*	*ERG24*	215140	[106]
CHILD syndrome	*NSDHL*	*ERG26*	308050	[107]

Additionally, HMGCR is the target of statins, one of the most widely-used classes of drugs used to treat cardiovascular disease and hypercholesterolemia by lowering circulating levels of low-density lipoprotein cholesterols[108–110]. Statins are competitive inhibitors that bind in the active site of HMGCR, slowing sterol biosynthesis by attacking the rate-limiting step[110]. However, all individuals are not equally sensitive to statins, a variation that is thought to be due to genetic variance among humans, including in the *HMGCR* gene itself, as well as in regulatory genes[111]. Ever a useful model for study of human biology and genetics, yeast has been used as an expression model to study the functional variability of HMGCR mutants, a model which could be extended to complement additional human genes in yeast and study their coordinated effects on statin sensitivity[112]. Since HMG-CoA reductase (*ScHMG1+ScHMG2/HsHMGCR*) is so tightly regulated, a yeast model where it is expressed at the native promoters could improve the accuracy and applicability of these models. While the utility of humanized yeast models when it comes to studying drug targets and complex human genetic disease is clear, the precise substitution of

coding sequences and introduction of exact mutations requires the use of novel and precise gene editing tools.

1.4 CRISPR-Cas9 gene editing in *S. cerevisiae*

The efficient homology-directed repair (HDR) machinery in *S. cerevisiae* makes it straightforward to incorporate exogenous DNA fragments into the genome, as long as the fragments contain two "homology arms". Classically, auxotrophic or antibiotic selection markers have been used to select for strains in which recombination was successful, but recently, nucleases that induce double-strand breaks (DSBs) at the target locus have been preferred to select for desired engineered strains[3]. DSBs are lethal if not repaired and, as *S. cerevisiae* prefer HDR over NHEJ, only cells that can incorporate the DNA repair template will survive, allowing marker-less selection[113]. Nuclease systems native to yeast, like I-SceI, are effective at introducing DSBs and integrating recombinant DNA in yeast, but only at their particular target sites and upon induction of the I-SceI nuclease, which limits their applications[114]. The *delitto perfetto* method improves upon this by including an initial step of integrating a template at any desired location in the genome and following that by transformation with the desired recombinant fragment, but this still requires two transformations for every engineering event[52,115]. The advent of the CRISPR-Cas9 endonuclease system has revolutionized yeast genome editing by enabling single-step targeting of any locus in the genome[116,117].

The Cas9 (CRISPR-associated protein 9) endonuclease originates as a bacterial defence system, similar to restriction endonucleases, but it is unique in that the target sequence is customisable and adaptable. A small guide RNA (sgRNA) sequence, which can be defined by any user designing the system, instructs the Cas9 nuclease where to create a double strand break with

precision. The only universal requirement for this system is the protospacer adjacent motif (PAM), a 5'-NGG-3' sequence that exists all over the genome, making this the most versatile strategy for genome editing[117]. When editing most human cell lines, which prefer NHEJ over HDR, additional steps are required to increase the efficiency of repairing these DSBs, but in *S. cerevisiae*, the integration of a recombinant repair template with homology arms is highly effective[113]. More than one sgRNA sequence can be expressed at once, allowing multiple loci to be targeted simultaneously[118].

Our lab and collaborators have previously designed a single-step, marker-less genome editing protocol for yeast, and applied it towards humanization of multiple genes at the native yeast locus[42,64]. Based on the Modular (MoClo) Yeast Toolkit, a single CRISPR-Cas9 expression plasmid with a user-defined 20bp small guide RNA (sgRNA) target sequence attached to the scaffold RNA is transformed into yeast along with a linear DNA repair template of the gene of interest with homology arms on either side[13]. Plasmid selection selects only for yeast successfully transformed with the CRISPR plasmid, and the requirement to repair Cas9-induced DSBs using the repair template selects for successfully edited yeast clones. This method is particularly effective when editing essential genes, as even less efficient and error-prone NHEJ repair, which would result in loss-of-function of the essential gene, will lead to lethality. However, when editing essential genes, the function of that gene must be preserved by whatever repair template is integrated – when the repair template is a replaceable human gene, the result is humanized yeast, in which the human gene is expressed under the same conditions as the native yeast gene while recovering its lost function. This system has been employed to humanize genes in the proteasome, in heme biosynthesis, in the cytoskeleton, and now, in the sterol biosynthesis pathway[38,42,67].

1.5 Study goals

In this study, I used the CRISPR-Cas9-based humanization strategy as previously described to sequentially humanize multiple genes in the sterol biosynthesis pathway in yeast. Previous studies have indicated this pathway is almost entirely replaceable except for two genes (Figure 4) – we hypothesize that by progressively humanizing the pathway and making the genetic interaction landscape more similar to that of a human cell, some of these genes may become replaceable. The cholesterol pathway is essential for health and proper function of cells, and proper regulation of the pathway is essential for human health. A humanized yeast with a fully replaced sterol biosynthesis pathway would have many applications: it could be used as a platform to screen novel drugs that target genes in this pathway; to elucidate the mechanisms between statin sensitivity and genetic variance; or to study multifactorial diseases like porokeratosis, if all the implicated genes were humanized and their phenotypes characterized in yeast. Multi-gene humanized strains in this pathway also have the potential to help unravel essential genetic interactions and pinpoint evolutionary divergence between the yeast and human pathways, such as why *HsFDPS* is the only gene not replaceable. (*HsFDFT1* has, since this figure was published, been found to functionally replace a loss of function temperature-sensitive allele of *ERG9* via plasmid-based complementation[41].)

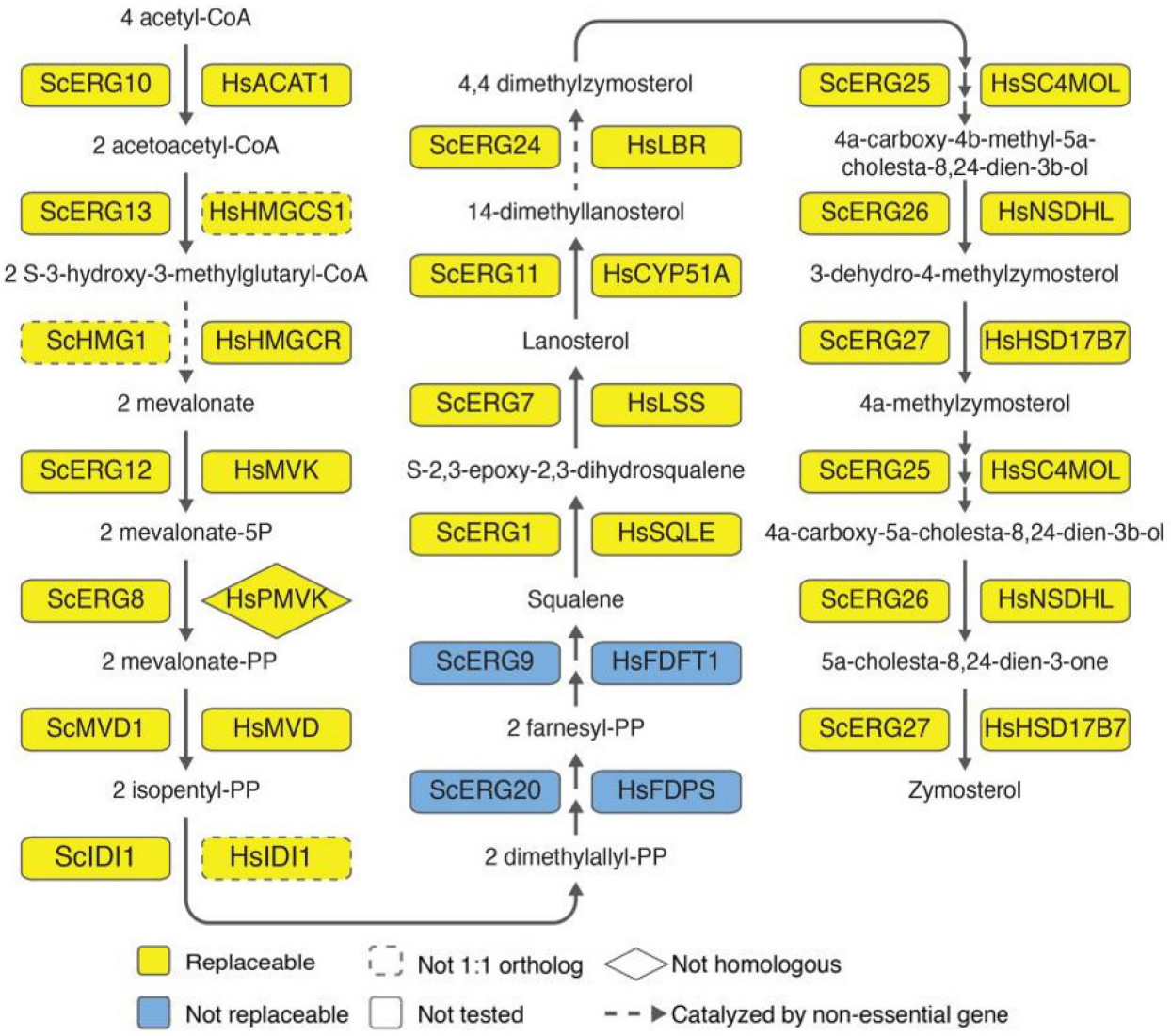


Figure 4: Sterol biosynthesis genes in yeast are functionally replaceable by their human orthologs

The sterol biosynthesis pathway is one of the most highly replaceable modules on a gene-by-gene scale, with fourteen of the sixteen upstream pathway genes functionally complementing either using temperature sensitive alleles, repressible promoters like a tetO promoter, or genetic segregation to select knockouts of the yeast gene when the human ortholog coding sequence is provided on a plasmid with constitutive or regulated expression. The yeast and human pathways diverge after the production of zymosterol to produceergosterol and cholesterol, respectively. *From Kachroo et al*[24].

This work was initiated at the University of Texas at Austin, where several single-gene humanized strains, as well as a 5-gene humanized strain assembled sequentially, were generated. In this thesis, I characterize and genotype these existing strains and generate novel ones, towards a fully humanized sterol biosynthesis pathway. Until now, few humanized pathways studied in

yeast have been essential, unlike sterol biosynthesis – this creates additional challenges. We hypothesize the editing of these genes will lead to altered fitness and phenotypes in yeast, which we will investigate to further understand the role of the pathway in cell function. Chapter 2 will describe the materials and methods used to engineer and validate novel strains, as well as methods used to generate and analyze data. Chapter 3 will describe my success at engineering novel humanized strains for the sterol biosynthesis pathway, both at the level of individual genes, as well as multiple genes of the pathway in a single strain, highlighting potential engineering bottlenecks. Chapter 4 will describe multiple methods for characterizing these strains, including their growth, protein expression, and protein localization. Chapter 5 will focus on future directions and strategies towards humanizing the remainder of this pathway in yeast, including preliminary results on a novel method towards engineering as yet non-replaceable human genes in yeast.

Chapter 2 – Materials and Methods

2.1 Gene-phenotype association data curation and code

In-depth methods (R markdown file) detailing data curation and sources for Figures 1 and 2 can be found on GitHub (https://mlvandeloo.github.io/kachrooyeasthumanizationreview/).

Raw data for yeast single genotype–phenotype associations and yeast-human gene complementation pairs were acquired from the Saccharomyces Genome Database (SGD) [6]. These data include gene deletion assays, cellular morphology data, and chemogenomic screens performed in yeast[7,25,119–121]. Human gene–phenotype associations were retrieved from OMIM (https://www.omim.org/)[122]. Additional OMIM data (OMIM gene IDs that have an associated phenotype with gene IDs) were retrieved from NCBI's FTP server (https://ftp.ncbi.nih.gov/gene/DATA/mim2gene_medgen). Orphanet data was obtained from Orphadata: free access data from Orphanet © INSERM 1999, available on https://www.orphadata.org, data version 1.3.15/4.1.7. Lists of yeast genes, their replaceable human orthologs and their essentiality status were obtained from SGD. The associated OMIM and Orphanet entries were generated based on data from the above-mentioned sources.

Figures 1-2, including the methods described above, are published in our review article[1].

2.2 Strains and media

All strains were grown in Yeast Peptone Dextrose rich medium (10g/L yeast extract, 20g/L peptone, 20g/L dextrose, +/- 20g/L agar for plates, all from Sigma Aldrich). For plate selection, synthetic drop-out medium was used, with synthetic drop-out media for uracil and leucine as required (20g/L dextrose, 20g/L agar, 1g/L glutamate, 2g/L yeast nitrogen base without amino acids, and 1.7g/L yeast synthetic dropout mix (uracil/leucine), all from Sigma Aldrich).

Wild-type strains used in this study were BY4741 (genotype: *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and BY4742 (genotype: *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) [123]. All humanized strains were engineered using either BY4741 or BY4742 as a starting platform, and were named systematically according to genotype, using the format HsSX.Y, where "Hs" represents the organism source of the engineered genes (Homo sapiens); "S" represents the pathway (Sterol biosynthesis); "X" represents the number of replaced genes in the strain; and "Y" represents an individual strain, with a new number for each strain generated. See Table 3 and Table 4 for a complete list of strains generated during this study.

2.3 Plasmid construction

All Golden Gate cloning assemblies of the CRISPR-Cas9 target (sgRNA) vectors were done in the pDirect-CEN6-URA-KanR-GFP-dropout-Cas9 plasmid derived from the MoClo collection[13]. For Golden Gate cloning assembly of the *LEU2* homology repair template plasmids, the integrating pYTK097 vector derived from the same collection was used. For all assemblies, 20fmol of plasmid DNA were used, to 40fmol of each insert (a single dsDNA sgRNA insert with complementary overhangs for CRISPR-Cas9 plasmid assembly, and three PCR-amplified inserts with cut sites and complementary overhangs added by PCR for *LEU2* locus homology plasmids). Type IIS restriction enzyme and buffer used for the reactions was FastDigest BsmBI (Thermo Fisher, cat. No. FERFD0454).

All Gateway cloning assemblies for GFP-tagging of human proteins in yeast expression vectors were done using LR and BP clonase mixes from Invitrogen (cat. No. 11791020, 11789100). Entry clones of human genes were either generated by Gateway cloning BP reactions using PCR-amplified fragments with attB1 and attB2 sites recombined into the donor vector

pDONR221 (Addgene ID 2394), or taken directly from the human ORFeome collection, as all clones in that collection are Gateway-compatible and without stop codons. The destination vector pAG416-GPD-ccdB-EGFP was paired with each respective entry clone in an LR reaction to generate the GFP-tagged human gene expression vectors. GFP-tagged human genes were under the control of the constitutive yeast GPD promoter for expression in yeast.

All plasmids were validated by restriction enzyme digestion and by Nanopore plasmid sequencing by Plasmidsaurus (https://www.plasmidsaurus.com/).

2.4 *E. coli* transformation

DH5α competent *E. coli* cells (NEB, cat. No. C2987H) were used for transformations. To one unit of cells (50µL), 1ug of plasmid DNA was added. The mixture was incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds, and re-placed on ice to recover for 5 minutes. 750µL of either LB broth or SOC media (NEB) was added, and suspensions were placed on a shaker at 37°C for at least 45 minutes. Cells were collected by centrifugation, suspended in ddH2O, and streaked on antibiotic-selective LB plates.

For plasmid cloning and miniprep, successfully transformed cells were grown overnight in 5mL of LB broth + selective antibiotic, and plamids were isolated using a Spin Miniprep Kit (Qiagen, cat. No. 27106X4).

2.5 CRISPR-Cas9-mediated humanization at the native yeast locus

Human gene coding sequences were amplified using templates from the human ORFeome collection, the Mammalian Gene Collection from Horizon Discovery, or as gBlocks synthesized by Integrated DNA Technologies[124,125]. Long 100bp oligonucleotides complementing 20-30bp of

the human gene of interest on either end, plus 80bp of the yeast homology region up- and down-stream of the gene were used as primers for template amplification. PCR was performed using Phusion High-Fidelity PCR Master Mix with HF Buffer (ThermoFisher Scientific, cat. No. F531L), per manufacturer's instructions, +2.5% DMSO. Template fragments were verified on a gel and purified using the QIAquick PCR purification kit (Qiagen, cat. No. 28104).

Yeast transformation was performed one of two ways: using the Frozen-EZ Yeast Transformation II Kit (ZymoResearch, cat. No. T2001) according to the manufacturer's instructions, or using the lithium acetate method [64]. For transformations, 1ug of plasmid DNA was used, and/or 5ug of linear DNA to ensure high transformation efficiency. For both methods, yeast was grown to an OD_{600} (optical density at 600nm wavelength) of 0.7-1.0, and then made chemically competent by either method. When using the ZymoResearch method, yeast was frozen at -80°C after addition of the Frozen-EZ Solution 2, and then recovered when ready to proceed.

For the lithium acetate method, yeast was pelleted, washed with ddH_2O, and suspended in 100mM LiOAc to incubate for 10 minutes at 30°C. Cells were pelleted and suspended in a mixture of 40% PEG_{4000}, 100mM LiOAc, and 0.3mg/mL boiled salmon sperm DNA, added sequentially, followed by the transformant plasmid and/or linear DNA. The transformation mixture was incubated for 30 minutes at 30°C, followed by 20 minutes at 42°C, and finally pelleted and suspended in 100µL of ddH_2O.

For both methods, final suspensions were plated on appropriate selective media and incubated at 30°C for 3-4 days. Plasmid curing for G418-resistance was carried out by repeated streaking on rich media; for *URA3* plasmid, repeated streaking on rich media was used, or 5-FOA counter-selection if necessary (adapted from Cold Springs Harbour Yeast Genetics course)[126]. As selection for the linear DNA template, including the *LEU2* gene and homology for the ectopic

chromosomal locus experiments required genomic uptake of the *LEU2* locus from the linear DNA template, no plasmid curing was required.

2.6 PCR genotyping of engineered loci

Rapid PCR confirmation of humanized loci were done by colony PCR, using Phire Green Hot Start II PCR Master Mix (ThermoFisher Scientific, cat. No. F126L), according to the manufacturer's instructions. Single yeast colonies were picked with a pipette tip, mixed in 20µL of water, and incubated at 98°C for 5 minutes. For a 10µL reaction, 2µL of the colony suspension was used as a template.

For high-throughput PCR confirmation of multiple loci, purified DNA was used as a template in Phusion High-Fidelity PCR Master Mix with HF Buffer (ThermoFisher Scientific, cat. No. F531L), per the manufacturer's instructions. For DNA extraction, yeast was grown in YPD overnight, pelleted, and suspended in 0.1M sorbitol, 0.1M Na_2EDTA pH 7.5, 0.5mg/mL Zymolyase 20T and incubated for 1 hour at 37°C. Cells were pelleted and suspended in 50mM Tris-HCl pH 7.4, 20mM Na_2EDTA, 1% SDS, and incubated for 30 minutes at 65°C, before adding up to 1.3M KOAc and placing on ice for 1 hour. After centrifuging and discarding the pellet, one volume of 100% isopropanol was added to the supernatant. Mixture sat at room temperature for 5 minutes before briefly centrifuging, pouring off the supernatant and air-drying the pellet. The pellet was suspended in TE buffer. 10ng of DNA were used per PCR reaction.

For certain larger numbers of PCR validation experiments, fragment analysis was performed using the 5400 Fragment Analyzer System by Agilent, supported by the Concordia Genome Foundry.

2.7 Sequencing confirmation of humanized loci

Sequences were amplified using Phusion High-Fidelity PCR Master Mix with HF Buffer (ThermoFisher Scientific, cat. No. F531L), and sequenced using primers designed at least 100bp up- and down-stream of the gene of interest. Fragments were fractionated on agarose gel to verify and isolate the correct bands, and then extracted and purified using QIAquick Gel Extraction Kit (Qiagen, Cat. No. 38706X4). All locus Sanger sequencing was performed by Eurofins Genomics.

2.8 Whole genome sequence analysis parameters

All engineered strains used in whole-genome sequencing (WGS) analysis were built by Mudabir Abdullah (humanized proteasome strains). Genomic DNA was purified from a wild-type yeast strain, a singly-humanized Hsalpha1 strain, and a quintuple-humanized Hsalpha1, alpha2, alpha3, alpha4, alpha5 strain.

Wild-type and engineered strains were sequenced using Illumina MiSeq 2x150 at 30x coverage using 150-bp paired end reads. The Geneious Pro Software and its included tools were used for pairing paired-end sequences, trimming ends and adapters based on quality using the BBDuk tool, and reference mapping using the Geneious Read Mapper algorithm at medium-low sensitivity, iterated 5 times[127]. Reads were mapped to a reference BY4741 strain (for the engineered strain, the same reference sequence was used, but replacing the sequences of the 5 engineered genes with their human ortholog to ensure alignment with the humanized loci reads) using the Geneious Prime Software 2020.2.4 algorithm. Mapping was run at medium sensitivity. Word length to allow for matching was set to 18 with a maximum permitted mismatch of 20% of the read length, maximum mismatches per read were set to 20%, a minimum 80% overlap was required, and reads with errors were set to accurately be mapped to repeat regions – this was

iterated 5 times to give the final mapping. For analysis, low coverage regions of below 2 standard deviations from the mean were excluded from SNP-calling, and only SNPs with a variant frequency of 0.90 or higher were considered. SNPs were called with a minimum variant frequency of 0.25, and low coverage regions of below 2 standard deviations from the mean were excluded. SNPs that were unique to the engineered strain (i.e., were not in the mappings of the wild-type strain) were marked. To ensure SNPs were not introduced by off-target CRISPR effects, 300bp up- and down-stream of each SNP was searched for Cas9-sgRNA alignment at 75% sequence similarity. Analysis and filtering were conducted using Microsoft Excel.

Figure S1 and the methods described above are published on bioRχiv[67].

2.9 Growth assays

Strains were seeded at an OD_{600} of 0.01 in 150µL of YPD and grown at 30°C for 72 hours in 96-well Corning plates, in triplicate for every strain, using 150µL YPD alone to blank. A Tecan Sunrise Microplate Reader running Gen5 software was used for measurement and raw data analysis. Data was analysed and figures were generated using R.

Spotting assays were performed by growing strains overnight, measuring the OD_{600}, then creating serial dilutions of each strain beginning at OD 1.0 and decreasing 10-fold with each dilution. A stamp was used to plate a small amount of each strain on rich media, in triplicate. Plates were left to grow for 4 days at either 25°C (room temperature), 30°C (yeast growth temperature), or 37°C (elevated temperature).

Growth rates and maximum doubling times were calculated based on the highest average growth rate/doubling time over a continuous three-hour period.

In-depth methods (R markdown file) detailing data analysis and sources for Figures 11 and 12 can be found on GitHub (https://mlvandeloo.github.io/MScthesisgrowth/).

2.10 LC/MS sample preparation and data analysis

Yeast whole cell lysate was prepared in a detergent-less lysis buffer as follows: 5mL of yeast grown to stationary phase were pelleted and suspended in approximately 5 volumes of lysis buffer (10mM TCEP, 40mM chloroacetamide, 8M Urea in 100mM Tris HCl pH 8.2) and vortexed for 1 minute. Lysate was sonicated 4 x 1 minute, with one-minute intervals on ice in between. Lysate was centrifuged at 4300g for 10 minutes, and supernatant was collected into a new tube. Samples were digested with trypsin and analysed by Liquid-Chromatography/Mass Spectrometry.

Raw mass spectroscopy data were read into the free Scaffold software view (Proteome Software, Inc., Portland, OR; version 5.1.2) and quantitative exclusive spectral counts were exported into R. Within Scaffold, peptide probability thresholds were set at a minimum of 95%, and a minimum of 80% for protein calls, with a minimum peptide requirement of 1. The edgeR package for differential gene expression analysis was used to calculate log fold changes and p-values across samples, as well as identify up- and down-regulated genes[128,129]. Enriched KEGG pathways were compiled using the clusterProfiler package for omics data[130]. The Complex Heatmaps package was used to generate and cluster all heatmaps[131,132].

In-depth methods (R markdown file) detailing data analysis and sources for Figures 15-18 can be found on GitHub (https://mlvandeloo.github.io/MScthesisLCMS/).

2.11 Microscopy

Brightfield and eGFP fluorescence images were taken using the Leica DMi6000B inverted epifluorescence microscope hosted at the Centre for Microscopy and Cellular Imaging at Concordia University. All imaging used the 63x bright lens (numerical aperture 1.4) with oil at 10x magnification. All images were captured using the Hamamatsu ORCA R2 C10600-10B digital ccd camera, and edited using ImageJ software.

2.12 Colony forming unit assays

Equal quantities of yeast were transformed with either an empty vector containing the appropriate selection, or with a CRISPR plasmid (plus linear template DNA, for CRISPR transformation experiments). Efficiencies of CRISPR experiments with various repair templates were calculated by dividing the number of colony-forming units (CFUs) observed in the experimental condition (CFU_O) by the number of colony-forming units observed from the empty vector transformation (CFU_E).

Chapter 3 – Humanizing the sterol biosynthesis pathway in yeast

3.1 Design and testing of CRISPR-Cas9 plasmids

In this study, I employed the previously-described CRISPR-Cas9-based strategy to engineer four new humanized strains with a single gene from the yeast sterol biosynthesis pathway replaced with its human ortholog[64]. Yeast with the wild-type locus for the gene to be replaced are transformed by i) a CRISPR-Cas9-sgRNA plasmid targeting the yeast gene for DSBs, and ii) a repair template consisting of the human ortholog coding sequence with homology to the native locus on either side that will be inserted by HDR (Figure 5). Using this strategy, I generated four novel strains, and also took four strains previously generated by collaborators at the University of Texas at Austin. I characterized the engineered loci in each of them using both PCR and sequencing. Most of the CRISPR-Cas9 reagents used were designed by previous collaborators, with the exception of the plasmids targeting *ScHMG2* and *ScMVD1*, which were designed by me, using the same steps described herein.

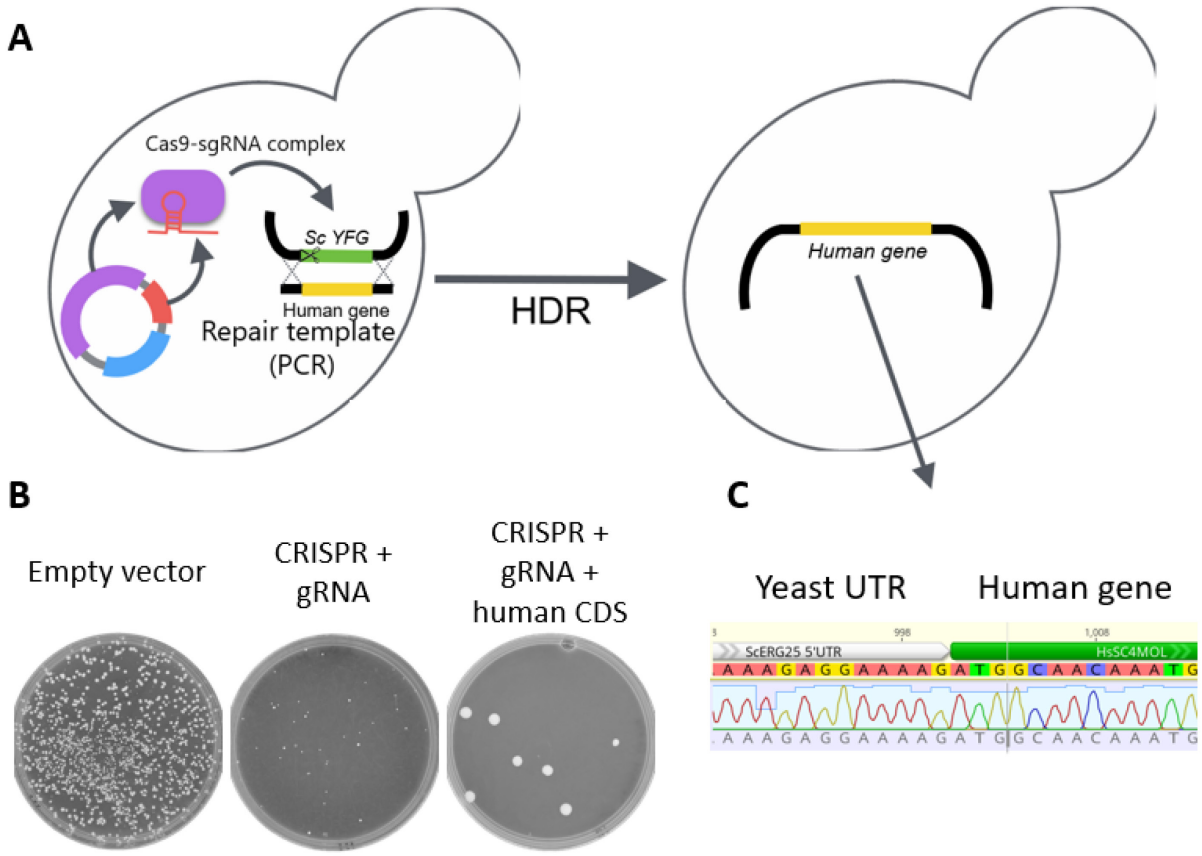


Figure 5: CRISPR-Cas9-based humanization approach in yeast

(A) A plasmid expressing CRISPR-Cas9 with a customized sgRNA sequence to target the desired yeast gene (*ScYFG*) and a repair template with the human gene ortholog and homology arms are transformed into non-engineered yeast. The Cas9-sgRNA complex induces DSBs in the yeast gene, which are repaired by HDR with the repair template, generating yeast with the human gene ortholog at the native locus. (B) While CRISPR-Cas9 on its own induces lethal DSBs, the addition of the human gene coding sequence (CDS) in the repair template will rescue the yeast if it functionally complements the deleted gene. (C) Sequencing of this engineered locus confirms insertion of the human gene (green bar) directly at the start codon (ATG) following the 5'UTR of yeast (yellow bar), as the Sanger sequencing read (pale blue) is continuous.

All CRISPR-Cas9 reagents used in this study were assembled using Golden Gate cloning, which allows directional and irreversible assembly of one or more DNA fragments into a vector. I used the pDirect-CEN6-URA-KanR-GFP-dropout-Cas9 vector as a backbone, with a unique sgRNA target sequence of 20bp specific to each yeast gene (Table 2)[64]. The 20bp sequences were designed to target the yeast gene in question, with 5' and 3' overhangs that align with those generated in the pDirect-GFP-dropout vector after digestion with type IIS restriction enzymes.

After transformation into *E. coli* and growth on a selective plate, GFP-negative colonies normally represent clones of properly-assembled plasmids, where the GFP cassette has been lost in favour of the designed sgRNA sequence. I validated selected colonies by restriction enzyme digestion and Sanger sequencing for the novel ScMVD targeting plasmid, in order to verify fidelity of the sgRNA sequence before use.

Table 2: CRISPR sgRNA target sequences expressed on Cas9 plasmids

Yeast gene target	sgRNA sequence	% Kill efficiency
ERG10	CTAGCACCAGAACAACCCAA	98.7%
ERG13	GGTGCCGCAAGACCAACCGG	98.1%
HMG1	CTACTGATGGAACGAAATGG	92.0%
HMG2	GGTCTCCTGTGTCTGATACA	97.5%
ERG12	TTAAAGTCTACTTTACCCAT	98.5%
ERG8	CATGCTGTASCCCATCCTTA	97.6%
MVD1	ACGCCAATTAAGAAAGGAAA	98.4%
IDI1	AATTCGTCATCAATACATAG	97.9%
ERG20	GATCTTACCGATCTGTTCTG	98.5%
ERG9	GGCAATGTAAAGATTCGTAA	97.7%
ERG1	GGTGAATTGATGCAACCAGG	98.0%
ERG7	GTAGCCGTAAACTATATCGC	98.6%
ERG11	TTACGGAACAAAGAGTGCAA	97.7%
ERG24	TGATGGATATCACTACAGAT	98.2%
ERG25	GCTCACCGTCTATTCCACTA	97.5%
ERG26	TATGACATAGTGAATGTTAA	97.2%
ERG27	ATTCTGCCTCAATTGACCAG	97.0%

DSBs induced by CRISPR in essential genes, such as all sixteen sterol biosynthesis genes, are lethal (kill efficiency was >99.9% for all). To test this collection of CRISPR-Cas9 plasmids for its ability to induce lethal DSBs in essential genes, I transformed wild-type yeast with each of these plasmids. All CRISPR-Cas9 plasmids induced lethality in the yeast, shown by a significant lethal growth phenotype in transformed strains (Figure 6), compared to the normal growth observed after transformation with a plasmid without the CRISPR-Cas9-sgRNA.

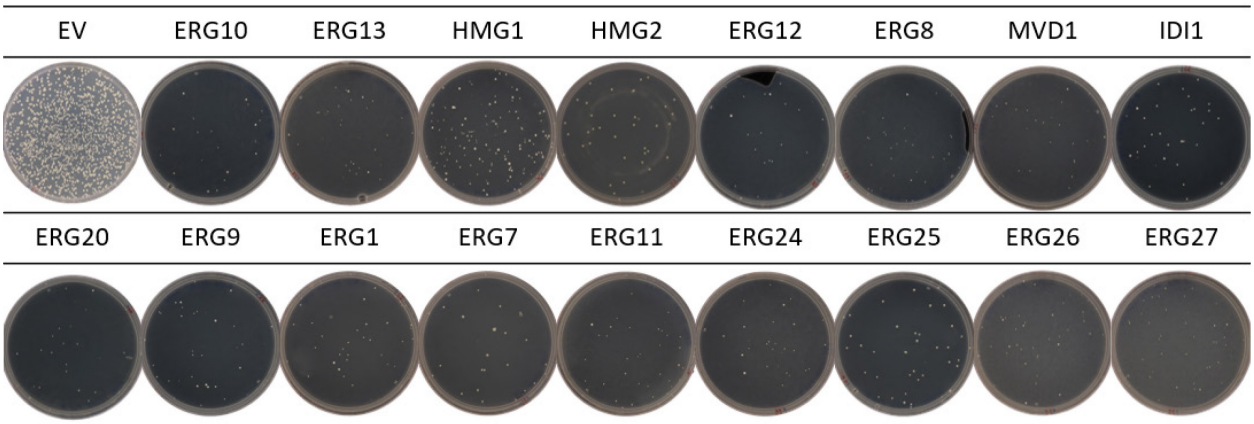


Figure 6: CRISPR-Cas9 plasmids induce lethal DSBs in essential yeast genes

Wild-type yeast were transformed with CRISPR-Cas9 plasmids without any repair template. Headings represent the native yeast sterol biosynthesis gene to which each sgRNA sequence was targeted. Compared to transformation with an empty vector (EV, no sgRNA sequence), DSBs introduced by the CRISPR plasmids were lethal to all yeast.

3.2 Exogenous PCR repair template for humanization

Humanization of the yeast sterol pathway genes requires a second element – a repair template containing the coding sequence of the human ortholog of the yeast gene to be replaced, including homology arms to the upstream and downstream regions of the corresponding yeast gene. Previous work has shown that a minimum homology length of 70bp on either end is sufficient for the humanization of most yeast genes[64]. For the sterol biosynthesis genes, homology arms of 80bp

were added onto the human gene coding sequence template by PCR, resulting in a single DNA fragment containing 80bp 5' homology, the human gene, and 80bp 3' homology[125].

If a yeast gene is functionally replaceable at the native locus, then co-transformation of wild-type yeast with both the CRISPR-Cas9 plasmid targeting a yeast gene and the repair template for the orthologous human gene should rescue the strain. The high efficiency of HDR-mediated DSB repair in yeast, when a complementing repair template is provided, allows the human ortholog to be inserted at the native yeast. As all genes in the sterol biosynthesis pathway are essential for growth, humanized strains will only grow if the human ortholog adequately complements the loss of yeast gene function. The four humanized strains already engineered by University of Texas colleagues were *ScHMG1/HsHMGCR, ScERG8/HsPMVK, ScERG9/HsSQLE*, and *ScERG24/HsLBR*. Therefore, this method was attempted for the remaining twelve of the sixteen sterol biosynthesis pathway genes and successfully generated four novel humanized strains for the gene ortholog pairs: *ScERG12/HsMVK, ScERG11/HsCYP51A1, ScERG25/HsSC4MOL*, and *ScERG26/HsNSDHL* (Table 3, Figure 7). I validated these by PCR and Sanger sequencing. The *ScIDI1/HsIDI1* strain exists, but was not available, and thus we could not characterize it further.

Interestingly, despite *ScHMG1/HsHMGCR* being the most highly regulated step of sterol biosynthesis, the humanized strain was easy to grow and characterize. *ScHMG2*, however, was not deleted or replaced in that strain, and efforts to induce a deletion at the *HMG2* locus or replace it with a copy of *HsHMGCR* were unsuccessful. This indicates the function and regulation of HMGCR in a humanized strain may be more complex – only once has *HsHMGCR* expression on a plasmid been demonstrated to complement the loss of function of HMG1 and HMG2 simultaneously[112]. Efforts continue to knock out or replace *HMG2* in the HsS1.1 strain – longer

growth recovery periods may be required, if the fitness defect is severe. Further experiments will also investigate if replacing *HMG2* with *HMGCR* instead of *HMG1* is possible.

Table 3: Humanized strains with a single sterol biosynthesis gene replaced to date

Name	Genotype
HsS1.1	*hmg1Δ::HsHMGCR*
HsS1.2	*erg12Δ::HsMVK*
HsS1.3	*erg8Δ::HsPMVK*
HsS1.4	*erg1Δ::HsSQLE*
HsS1.5	*erg11Δ::HsCYP51A1*
HsS1.6	*erg24Δ::HsLBR*
HsS1.7	*erg25Δ::HsSC4MOL*
HsS1.8	*erg26Δ::HsNSDHL*

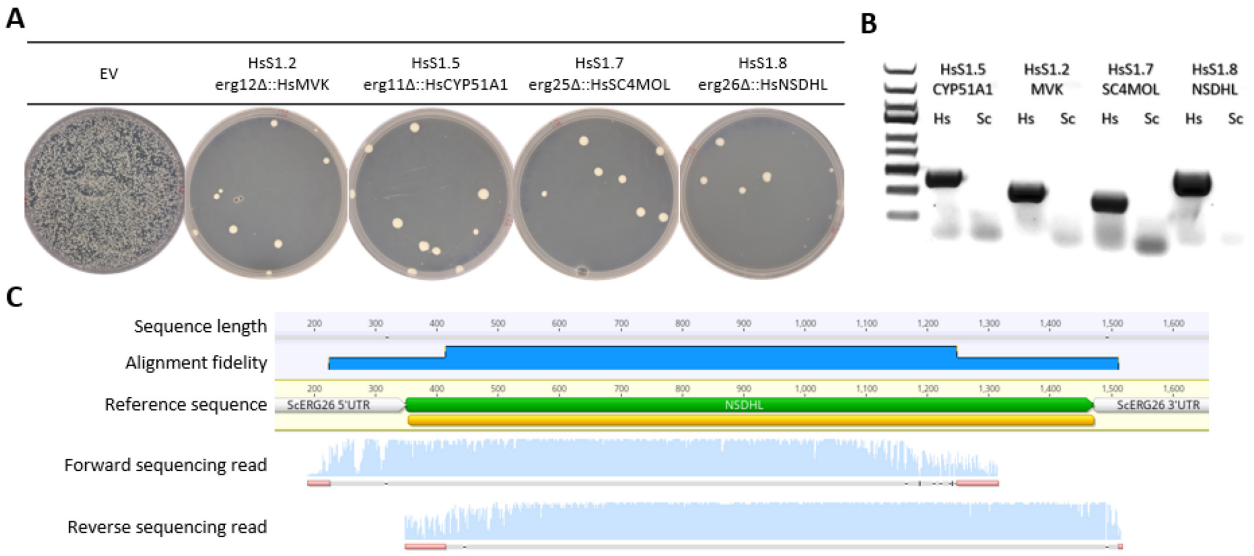


Figure 7: Novel individually humanized genes at the native locus in the sterol biosynthesis pathway in yeast

(A) After transformation of wild-type yeast with CRISPR plasmids targeting the yeast gene and a repair template for the human ortholog, four novel strains were created (genotypes listed in heading) – colony growth indicates functional complementation of the human gene inserted at the native locus. (B) Colony PCRs screens for presence of the human gene (Hs) and absence of the yeast gene (Sc) at the native locus, using a common forward primer outside of the homology region, and reverse primers targeted to each sequence. (C) Sanger sequencing of the engineered locus in strain HsS1.8 shows coverage of the 5'UTR and 3'UTR of the yeast gene (white bars) with the *HsNSDHL* sequence replacing the native *ScERG26* (green and orange bars). The bright blue band represents overlapped and aligned sequence coverage between the two Sanger sequence reads (pale blue).

Once strains were streaked out, I used PCR and Sanger sequencing to investigate newly-grown colonies for the presence of the human gene and absence of the yeast gene. Colony PCR was performed using a common forward primer that is specific to each locus (outside of the homology arm sequence) and a locus-specific reverse primer (one aligned with the human gene sequence and the other with the yeast gene sequence). Successful strains show amplification of the human-specific locus but not the yeast-specific locus, and the sequence aligns with the upstream promoter region of the yeast gene, followed by the human gene coding sequence as of the start codon, and the yeast terminator region following the stop codon.

The CRISPR plasmid paired with PCR-amplified human gene coding sequence strategy was only successful in humanizing eight out of sixteen of the genes in the sterol biosynthesis pathway. One possible reason for the inability to humanize these strains despite several attempts may be the low efficiency of the technique or less efficient HDR at these loci. Even when transforming yeast with a single product (such as a plasmid), only a small portion (<1%) of the competent cells used will successfully uptake and express the Cas9-sgRNAs. In the humanization strategy described herein, cells must successfully uptake two separate DNA molecules – the CRISPR-Cas9 plasmid and the linear repair template, lowering the percentage of cells recovered from the process significantly. In the future, for the set of replaceable sterol biosynthesis genes, a relatively high-efficiency method such as MERGE will be used to screen definitively for replaceability of these genes at the native loci [67]. The question of how to replace non-replaceable genes, as well as those that might not be replaceable at the native locus, remains, but may be addressed by our hypothesis that strains that are humanized to a higher degree will be more suitable platforms for replacing otherwise individually non-replaceable genes.

3.3 The ScEDIT toolkit for CRISPR-Cas9-based gene editing in *S. cerevisiae*

The CRISPR-Cas9 strategy of inducing lethality by targeting essential yeast genes with unique and customizable sgRNA sequences enabled the construction of a library of these highly-specific gene editing plasmids. In addition to the sterol biosynthesis pathway, other members of this and collaborating labs have created many CRISPR-Cas9-sgRNA plasmids for genes belonging to other critical pathways and complexes in yeast: The proteasome, including the alpha and beta core, the base, and the lid; the heme biosynthesis pathway; the cytoskeleton; the TRAPP complex; the glycosylation pathway; and a number of other isolated genes, as well as safe harbour sites

(intergenic regions which can be targeted for insertion of novel genes for metabolic engineering without any observable fitness cost)[21]. This library, referred to as ScEDIT, comprises >50 plasmids for the applications in editing yeast genes by inducing on-target lethal DSBs in the specified genes. We performed tests of a number of plasmids from this library, showing that all the included reagents induce lethal DSBs in wild-type yeast (Figure 8A). A number of the plasmids showed some leaky targeting, meaning that certain cells escaped CRISPR-Cas9-induced DSBs and continued to grow, but still at a significantly reduced rate when compared to the growth following empty vector transformation.

In addition to lethality induced by these CRISPR-Cas9 plasmids, humanized strains proved resistant to transformation with the plasmid that would have targeted the yeast ortholog now replaced with a human gene (Figure 8B). All of these humanized strains have previously been validated by both locus-specific and Sanger sequencing of the entire locus by other lab members. This resistance to lethal growth defects further demonstrates that once the yeast genes have been deleted from the genome of these strains, engineering to replace them with a functional human gene complement can take place, resulting in humanized strains, though the CRISPR reagents can be used for other gene knock-out applications as well by the yeast genetics community.

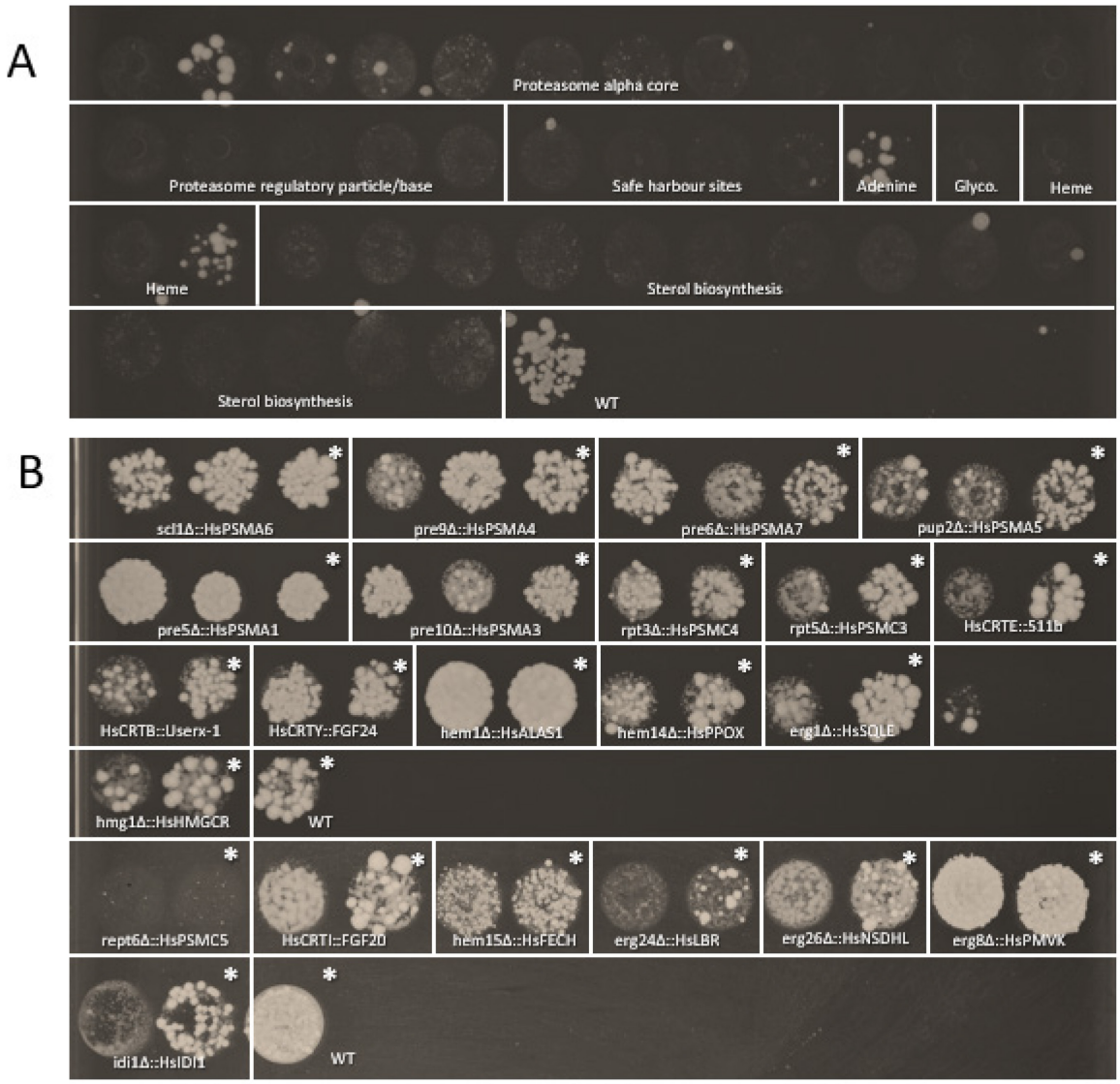


Figure 8: The ScEDIT collection of CRISPR-Cas9 plasmids for yeast genome editing

(A) A collection of CRISPR plasmids with sgRNA sequences targeting native yeast genes in multiple complexes and pathways induce lethal DSBs when transformed in wild-type yeast, compared to empty vector transformation (WT). Glyco. = Glycolysis, Heme = Heme Biosynthesis. (B) When the same CRISPR plasmids are transformed into humanised strains where the human gene has replaced the native yeast ortholog, strains remain viable compared to empty vector transformation (*) because they lack the unique sgRNA target site.

Robust tools and methods such as these represent a powerful addition to an extensive library of genetic manipulation tools for yeast. The CRISPR method has been demonstrated to be a reliable way to humanize individual genes at the native yeast locus, including multiple genes in the sterol biosynthesis pathway. Among these humanized strains created in my project and by collaborators

are strains with functionally complementing, natively expressed human genes that are critical targets for drugs, and that have associations with known human diseases. Humanized yeast will serve as a powerful platform of study for genetic interactions and human disease. However, there are limits to what can be studied with only a single altered gene, as all genes interact in complex networks. Furthermore, genes that are not replaceable may be so due to lack of critical interactions between the human protein and other proteins in a hybrid state. The next logical step is to take single-gene humanized yeast further, to humanizing entire pathways in yeast.

3.4 Engineering pathways in yeast

Previous work has provided strong evidence that genetic modularity is highly predictive of replaceability in a humanized yeast context – that is to say, genes found in pathways or complexes with many other replaceable genes are likely to be replaceable themselves[24]. The genes in the sterol biosynthesis pathway were highlighted as mostly replaceable on an individual level, with fourteen out of the total sixteen demonstrated to be replaceable using plasmid-based complementation. A sequential humanization strategy, where each yeast gene is replaced one-by-one by its corresponding human ortholog, uses the previously humanized strain as a platform to replace the next gene. I generated 5 novel multi-gene humanized strains, towards the ultimate goal of replacing all sixteen genes in this pathway (Table 4). Each gene was engineered using the CRISPR-Cas9 strategy as described earlier.

Table 4: Humanized strains with multiple sterol biosynthesis genes generated to date

Yeast gene deleted	hmg1Δ	erg12Δ	erg8Δ	idi1Δ	erg9Δ	erg1Δ	erg11Δ	erg24Δ	erg25Δ	erg26Δ
Human ortholog inserted	**HMGCR**	**MVK**	**PMVK**	**IDI1**	**FDFT1**	**SQLE**	**CYP51A1**	**LBR**	**SC4MOL**	**NSDHL**
HsS5.1[1]	●		●	●		●		●		
HsS6.1	●		●	●		●		●	●	
HsS7.1	●		●	●		●		●	●	●
HsS8.1	●		●	●	●	●		●	●	●
HsS9.1	●	●	●	●	●	●		●	●	●
HsS9.2	●		●	●	●	●	●	●	●	●

1 HsS notation refers to the strain type: Homo sapiens (Hs) Sterol (S).

A strain already harbouring five human genes, all of which were shown to be individually replaceable – HsS5.1, with the replaced genes *ScHMG1/HsHMGCR, ScERG8/HsPMVK, ScIDI1/HsIDI1, ScERG1/HsSQLE*, and *ScERG24/HsLBR* – was generated by previous lab members. While this strain was also generated by sequential humanization, the intermediate strains (with 2, 3 and 4 genes) were not available for further study. I confirmed this strain, as well as the subsequently generated ones, by locus-specific PCR and Sanger sequencing of all five loci, as previously described. For each novel gene integration, transformation-competent cells were made from a humanized strain, and the CRISPR-Cas9 and repair template method was tried for every remaining gene to be humanized. For each attempt, one gene of all those attempted was successfully replaced and used as a basis for the following humanization attempts, except for the addition of a ninth gene, where two different gene replacements were possible (Figure 9A). I

screened colonies for the presence of the human gene and the absence of the yeast gene by colony PCR and Sanger sequencing (Figure 9B-C). For the addition of the ninth gene, two strains were generated: one with *HsMVK*, and one with *HsCYP51A1*, which provides a versatile platform for further humanization experiments, as different "routes" of humanization may become clear from these two strains, enabling comparison of why certain genes are replaceable or not at certain stages. Work at generating a 10-gene strain containing both *HsMVK* and *HsCYP51A1* is ongoing, as is work towards using these strains as platforms in which to humanize the remaining six genes.

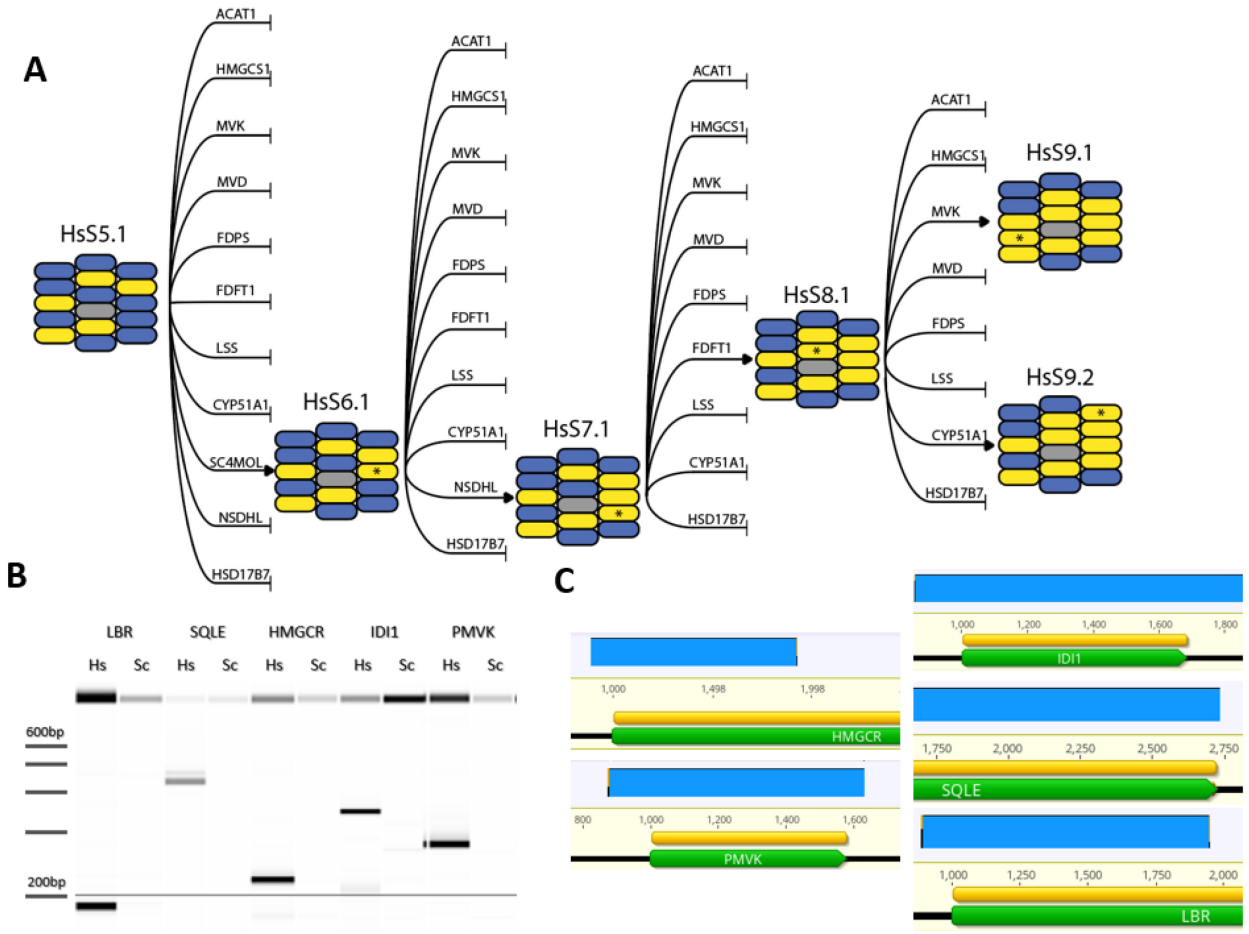


Figure 9: Sequential humanization of multiple genes in the sterol biosynthesis pathway in a single yeast strain

(A) Illustrated pathway of sequentially humanized genes added in this study. Each human gene (written above the black arrows) repair template was transformed into the strain paired with the CRISPR plasmid targeting the yeast orthologous gene, and successfully validated replacements are illustrated, one at each step, and two in the final step, yielding 6 total strains. Grey boxes represent *ERG20*, the non-replaceable

gene; blue boxes represent native yeast genes in the sterol biosynthesis pathway; yellow boxes represent replaced genes; * represent the location of the gene in the pathway replaced at that stage (gene name can be found on the arrow line). (B) Example of colony PCR screens for presence of the human genes (Hs) and absence of yeast genes (Sc) in strain HsS5.1 at the native loci, using a common forward primer outside of the homology region, and reverse primers tailored to each sequence. (C) Sanger sequencing of the engineered loci in strain HsS5.1, with human genes inserted at the native yeast ortholog loci. Blue blocks represent where Sanger sequencing reads aligned with the reference sequence; green/orange blocks denote the human gene coding sequence inserted at the native locus. This sequencing strategy was repeated for all additional strains.

3.5 Partially humanized sterol pathways enable the replacement of individually non-replaceable genes at the native locus

The successful generation of the eighth gene strain was particularly interesting for the gene that was replaced. Based on previous studies on multiple complementation assays, the yeast copy of Farnesyl-Diphosphate Farnesyltransferase 1 (*ScERG9/HsFDFT1*) was only replaceable once under strict conditions, and it has never been successfully humanized at the native locus to date. ScERG9/HsFDFT1 condenses 2 farnesyl pyrophosphate molecules in a two-step catalysis process into a single molecule, squalene, in both yeast and human sterol biosynthesis, and is the first gene in the pathway to be involved in sterol biosynthesis exclusively. Additional investigations will be required to unravel which precise changes in the yeast strains made *ScERG9/HsFDFT1* replaceable at the native locus only following the humanization of seven prior genes. FDFT1 may have essential interactions with some of those seven human genes in this pathway that have yet to be fully characterized, but that are revealed by partially humanizing the entire pathway.

Repeated rounds of CRISPR-Cas9 plasmid transformations run the risk of introducing off-target DSBs and mutations. While selected sgRNA sequences were screened using BLAST search to ensure they targeted the single desired site and had low complementarity to sites elsewhere in the genome, *in vivo* targeting may be different. Sanger sequencing of the engineered locus can confirm the presence of the gene of interest, and identify mutations introduced into the gene or

flanking sequences, but does not rule out mutations elsewhere in the genome. Whole genome sequencing (WGS), on the other hand, can screen the entire genome of engineered, multi-gene humanized strains for both correct replacement of desired genes, and for off-target CRISPR effects. In the future, it could also be used to screen for potential suppressor mutations in other genes that are responsible for modified phenotypes, such as recovered growth in a slow-growing engineered strain.

Whole genome sequencing on the sterol biosynthesis strains remains to be done, but has been done previously on novel strains engineered by other members of the lab, towards humanizing the alpha proteasome core in yeast. The analysis of these strains confirmed multiple items. First, that every humanized gene was in fact integrated wholly at the native locus for the yeast ortholog, with proper start and stop codons and good sequence coverage, with no nearby mutations (Figure S1). Second, that while there were a number of SNPs identified by the filters, none were suspected to be due to off-target CRISPR effects. 300bp up- and down-stream of every SNP was screened for all CRISPR sgRNA sequence alignment, at 75% similarity, and no potential cut sites were identified (Table S4). We expect similarly low off-target CRISPR effects to be seen when the humanized sterol strains are eventually sequenced.

3.6 A novel, higher-efficiency strategy for gene humanization

My project has demonstrated that sequential humanization of genes in a single pathway using CRISPR-Cas9 is possible, and that the process yields irreversibly engineered strains with the human genes expressed at the native loci (9 of 16 loci in a single strain). However, a number of individual or combinations of humanized sterol genes were not permitted even for the human genes that are functionally replaceable in plasmid-based assays. This strategy shows promise for

continuing to humanize additional genes sequentially, but novel strategies are needed to address the remaining non-replaceable genes at the native locus. Increasing homology arm length from 80bp up to several hundred base-pairs was attempted, but efficiency remained the same.

Sequential humanization of entire pathways via the CRISPR-Cas9 method has its drawbacks. As the number of genes in a pathway increases, so too does the number of intermediate combinations of genes increase exponentially, each with its own fitness, growth rate, regulatory challenges, and tolerance for further rounds of CRISPR-Cas9 transformation and HDR. To overcome these drawbacks, the MERGE method was developed by our lab to perform combinatorial engineering of multi-gene pathways in yeast, with much higher efficiency and clear readout of viability of all possible combinations[67]. By leveraging the highly effective HDR mechanisms of yeast and mating yeast with different heterozygous combinations of engineered genotypes followed by Cas9-sgRNA-mediated gene drives, multiple humanized loci can be easily combined to create multi-gene strains. Given the efficiency of this strategy, we will attempt to create humanized alleles for the sterol pathway genes, and used CRISPR-Cas9 to convert the loci to homozygous humanized alleles to systematically determine which genes are replaceable at the native locus, and which are not.

While this strategy is highly efficient and accurate, as well as applicable for both single- and multi-gene humanization projects, it requires engineering at the native loci, which can be challenging. While five of the six remaining genes to be engineered in the pathway have yet to be humanized individually, there is a chance that they may become replaceable once a higher fraction of the pathway is humanized, in the same way that *FDFT1* was only replaceable in a strain with multiple humanized genes. From a mathematical standpoint, the low efficiency of the CRISPR method derives from the fact that two discrete DNA molecules (the CRISPR plasmid and the repair

template) must be transformed into the same cell simultaneously, and both must function as required (the CRISPR protein must correctly target the yeast gene, and the repair template must be available for HDR). The efficiency of MERGE is much higher, as it relies only on transformation of the CRISPR plasmid, while the chromosome-scale repair template serves as a highly efficient repair template. However, the strategy only works in the diploid context. By providing a stable endogenous repair template with 200-300bp homology inserted at non-homologous sites in the chromosome, this low efficiency may be overcome.

Towards that, I built a *LEU2* integrating plasmids containing an insert with 300bp homology to the 5′UTR and 3′UTR of the yeast gene of interest, flanking the human ortholog coding sequence. I tested this strategy using with *HsSC4MOL* – replaceable for its ortholog *ERG25* at the native locus, both individually and in a multi-gene pathway – and *HSD17B7* – the final in-common step of ergosterol and cholesterol biosynthesis, which has only ever been functionally replaceable for its ortholog *ERG27* by plasmid complementation[133]. After integrating the insert at the *LEU2* locus, I transformed the strains with CRISPR-Cas9 plasmids targeting *ScERG26* and *ScERG27* respectively. As expected, DSB repair for *ScERG25* was significantly improved by the non-homologous *HsSC4MOL* repair template. Comparatively, the efficiency of repair for the chromosomal *HsHSD17B7*-based DSB repair remained low, suggesting that this gene truly is not replaceable at the native yeast locus (Figure 20).

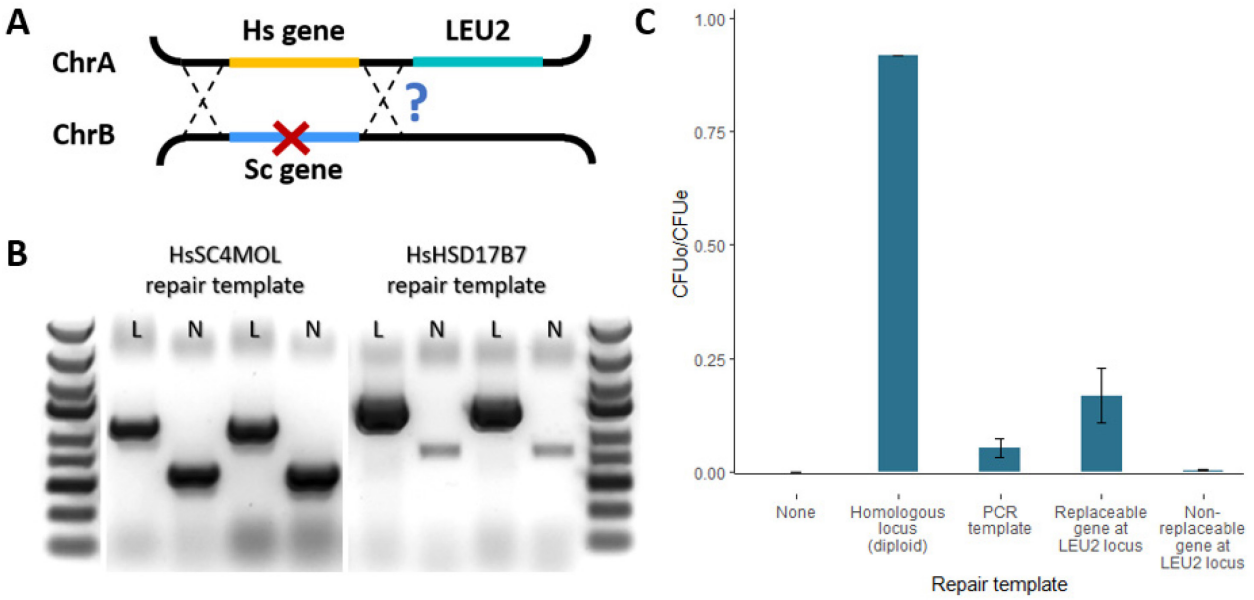


Figure 10: Ectopic chromosomal repair template increases HDR efficiency over exogenous PCR repair template for replaceable genes

(A) Illustration of HDR from a human gene (Hs) repair template inserted at an ectopic chromosomal locus (upstream of *LEU2*) with homology arms to the yeast native ortholog (Sc) after DSB (X) induced by CRISPR-Cas9. (B) Presence of the human gene repair template at the *LEU2* locus (L) or native yeast locus (N) (ChrVII for *ScERG25/HsSC4MOL*; ChrXII for *ScERG27/HsHSD17B7*) detected by PCR in two distinct colonies, indicating yeast can use ectopic chromosomal repair templates to repair DSBs. (C) Efficiency, expressed as CFU_O/CFU_E, of HDR following CRISPR-induced DSBs using different repair templates. Diploid heterozygous strains can use the viable homologous allele to convert the strain to a homozygous genotype with high efficiency; an exogenous PCR template provides low efficiency of HDR; insertion of a repair template of a replaceable gene (*HsSC4MOL*) with homology arms (to *ScERG25*) at the *LEU2* locus has a much higher repair efficiency than the same technique with a non-replaceable gene (*HsHSD17B7* with *ScERG27* homology arms).

Sequencing of the native locus to validate proper replacement of the gene is currently underway, though initial PCR-based confirmations showed successful humanization at the native locus for both genes (Figure 20B). These initial results are promising towards improving the efficiency of the CRISPR-Cas9 method by providing a stable endogenous repair template. Efforts at multi-copy gene editing using a similar CRISPR-Cas9 strategy coupled with HDR using a template integrated at an ectopic chromosomal locus have demonstrated that the length of recovery time in liquid media extends expression and activity of the CRISPR-Cas9 plasmid and leads to

increased rate of repair. In the future, this strategy could be applied to improve the efficiency of repair at the native locus[21].

While ten of the sixteen genes have been shown to be replaceable in at least one multi-gene context, six genes (one which has never been observed as replaceable – *ScERG20/HsFDPS* – and five that have successfully complemented in previous assays) have not yet been shown to be replaceable at the native locus, in either individually humanized strains, or multi-gene humanized strains. I hope to begin to elucidate the reasons behind some human genes becoming replaceable once neighbouring human gene contacts are restored, while other replacements remain elusive, by characterizing the expression and localization of these proteins in Chapter 4.

Chapter 4 – Characterization of humanized strains to identify pathway engineering bottlenecks

4.1 Utility of characterizing strains

Having generated multiple humanized strains of yeast, with single genes or multiple genes replaced, proper characterisation of these novel strains was paramount. This is necessary for several reasons. First, if these strains are to be used as a platform to investigate genetic interactions and human genetic diseases, then the baseline behavior of the strain needs to be well-known and measurable. For instance, if one were to introduce targeted mutations into a human gene in one of these yeast strains to investigate the effect of the mutation on growth as a proxy for how the mutation may be affecting human cells, knowing the baseline behaviour of the cell, such as its growth rate and gene expression profiles, is essential. The second reason for characterizing these strains has to do with modularity of replaceability of genes that were not possible to replace individually. Studying the conditions that allowed certain genes to complement the missing yeast orthologs may reveal clues as to why certain genes were replaceable, while others were not, indicating engineering bottlenecks in this pathway to be taken into consideration for future humanization projects.

For instance, of particular interest was strain HsS8.1, the first occurrence of replaceability of *HsFDFT1* at the native locus in yeast, where it was unable to be replaced individually at the native locus. More in-depth analysis of the HsS7.1 strain into which this gene was humanizable, and the HsS8.1 strain in which it is present, may reveal changes in conditions in the cell that enabled replaceability. The only unreplaceable gene in the sterol biosynthesis pathway remains *ScERG20*, whose human ortholog *HsFDPS* does not complement its lost function when expressed at the native locus or on a plasmid under constitutive promoter control, despite the fact that they

retain the same catalytic function. Five additional genes – *ScERG10/HsACAT1, ScERG13/HsHMGCS1, ScMVD1/HsMVD, ScERG7/HsLSS*, and *ScERG27/HsHSD17B7* – do not appear to be replaceable at the native locus, despite being replaceable when expressed on a plasmid under a constitutive yeast promoter. If qualities such as modified growth rates, changes in protein expression, novel genetic interactions, or differences of cellular localization of human and yeast proteins can be identified, it may indicate why certain genes remain non-replaceable at the native locus, and reveal what next steps should be taken to engineer them. These results will also have applications to improve efforts at humanizing genes and pathways other than the ones described herein.

4.2 Growth behaviour of humanized sterol strains

For humanized strains to become study platforms for screening drugs and investigating the impact of disease mutations on phenotype, the growth of these strains needs to be benchmarked. I measured growth curves of all strains over a 72-hour period, and observed several strains with significantly slow growth. Growth measurements, by quantifying the area under the growth curve, show this variety clearly, such as in the case of HsS1.1 (with a humanized *HMGCR*) that grows well, compared to the HsS1.2 (with a humanized *MVK*), which has a severe defect (Figure 11A). As a baseline, the wild-type strain reaches a stationary phase of approximately OD 2.0 after only 16 hours (Figure 11B). Strains HsS1.1 and HsS1.4 grow at nearly the same rate, reaching stationary phase after about 16 hours, but have lower peak OD. Conversely, strains like HsS1.2 and HsS1.8 take at least 40 hours to approach leveling off of the log phase, and have growth rates 0.20 and 0.35 that of the wild-type, respectively (Table 5). However, HsS1.2 reaches a much lower final OD. The remaining strains show a mix of reduced growth rates and lower cell density. These

results indicate that growth defects in humanized strains are specific to the particular replaced gene, and that these replacements may have multiple consequences for the cells revealed through these growth defects.

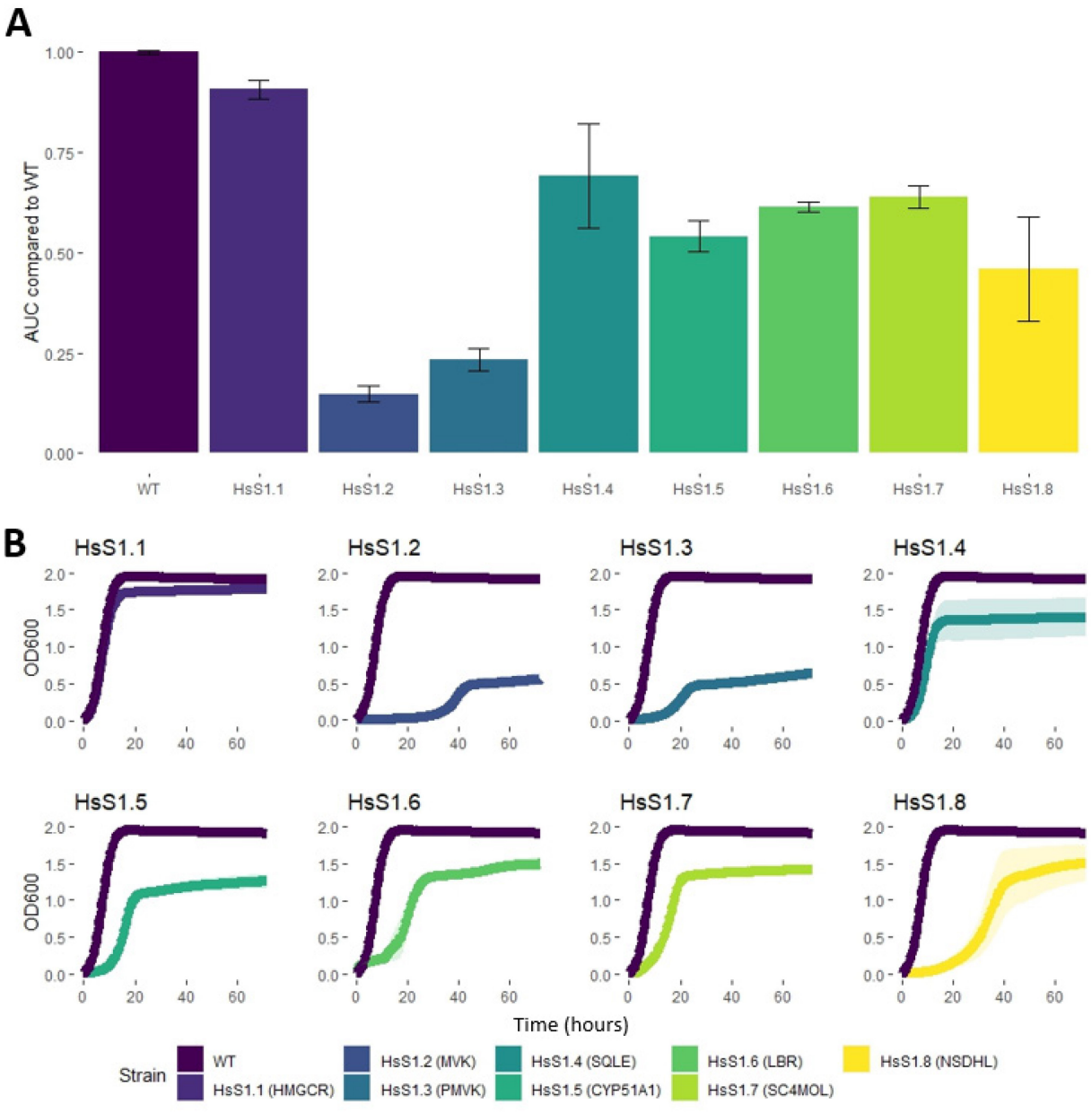


Figure 11: Growth behaviour of single-gene humanized strains

(A) Area under the curve (AUC) was calculated based on 72-hour growth curves conducted in triplicate (N = 3) for each engineered strain. AUCs represent a numerical value of yeast grown over this period. Error bars represent standard error between biological replicates. (B) Illustrated OD_{600} curves over a 72-

hour period conducted in triplicate. Shaded areas represent standard error between biological replicates. The darkest curve represents the wild-type (WT) curve for comparison.

The generation of multi-gene humanized strains shows an interesting path of fitness costs (growth defects) with progressive humanization. The strains show significantly reduced growth yields compared to the wild-type, but variability between sequentially engineered strains (Figure 12A). For example, beginning with strain HsS5.1, there is a clear growth defect, with the strain having a growth rate of 0.21 that of the wild-type, and only reaching a maximum cell density of approximately OD 0.7, compared to the OD 2.0 of the wild-type strain (Figure 12B, Table 5). However, the growth defects do not appear to be cumulative in multi-gene humanized strains, as the five singly humanized strains sharing genes with strain HsS5.1 have varied growth defects – HsS1.1, with a human copy of *HMGCR*, shows little to no growth defect, compared to HsS1.3 with a human copy of *PMVK*, which has a similar growth profile to HsS5.1. This suggests that growth rates of the multi-gene humanized strains may be limited by the growth rates of the least fit single genes. After I replace the sixth (*SC4MOL*) and seventh (*NSDHL*) genes at the native loci, however, the fitness of the strains continues to deteriorate, with strain HsS7.1 showing extremely reduced growth.

So far, higher degrees of humanization have led to progressively reduced fitness. However, *in silico* modelling has demonstrated that as the fraction of humanized genes in a pathway or complex increases, there are two likely possibilities for the relative fitness of the strains (Figure 13)[134]. A multiplicative model suggests that as more humanizations are performed, there is a progressively deteriorating effect on the fitness until the strain is no longer viable. Alternatively, a cooperative model suggests that the humanization of additional genes will eventually have a positive effect on the overall fitness. The initial few humanizations will show reduced fitness, followed by a recovery as the fraction of humanized genes begins to surpass the fraction of

remaining yeast genes. The hypothesis underlying the cooperative pathway is that there may exist essential species-specific genetic, physical or metabolic interactions between yeast and human genes, and that these interactions are disrupted in hybrid strains with a mix of human and yeast genes. As an entire pathway becomes more humanized, and interactions between multiple human genes begin to compensate for disrupted interactions between yeast genes, the fitness of the strain may benefit.

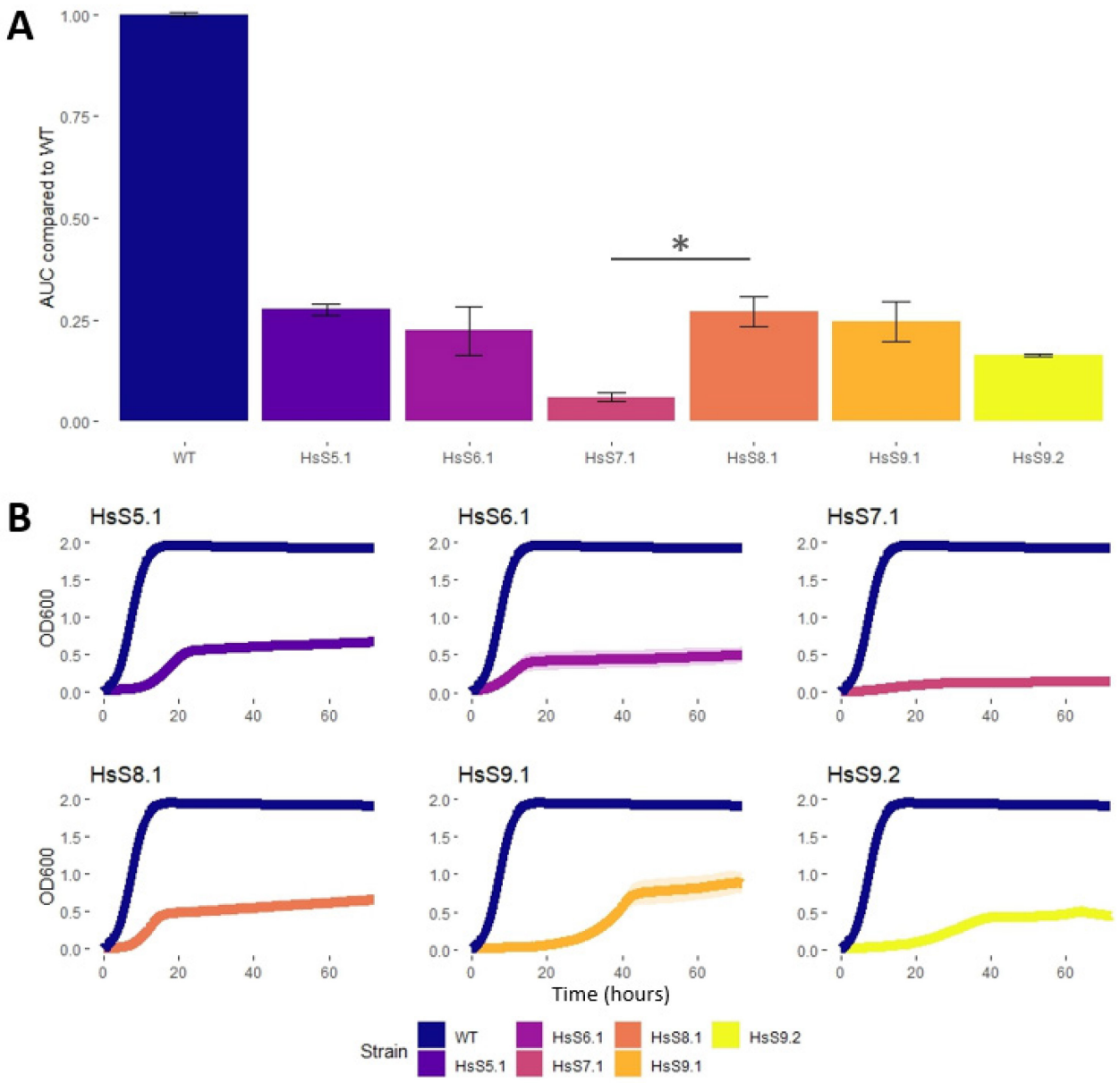


Figure 12: Growth behaviour of multi-gene humanized strains

(A) Area-under-the-curve (AUC) was calculated based on 72-hour growth curves conducted in triplicate (N = 3) for each engineered strain. AUCs represent a numerical value of yeast grown over this period. Error bars represent standard error between biological replicates. * = p-value < 0.05 (B) Illustrated OD_{600} curves over 72-hour period conducted in triplicate. Shaded areas represent standard error between biological replicates. The darkest curve represents the wild-type (WT) curve for comparison.

Strain HsS8.1 is of interest for two reasons. First, as discussed in Chapter 3, the eighth gene to be replaced was *ScERG9/HsFDFT1*, a gene that was not replaceable at the individual level at the native locus.[41] Second, the growth rate increased, both with a maximum cell density and growth

rate higher than that of HsS5.1 (Table 5). With the addition of a ninth gene (*ScERG12/HsMVK*), the overall maximum cell density remained high, but the prolonged lag phase as seen in HsS1.2 persisted. This result suggests that there is a significant regulatory change when yeast Mevalonate Kinase is replaced by its human counterpart. Nevertheless, the overall higher cell density observed in HsS9.1 suggests cooperative behaviour as the growth defects in the intermediate humanized strains are rescued for subsequent humanizations. Furthermore, while we can begin to associate the aberrant growth rates with certain individual genes (i.e., *MVK*), the possibility of rescuing growth defects in higher order humanizations suggests the effect of compounding multiple humanized genes together in a single strain is complex and unique to the genes involved.

Table 5: Growth parameters of humanized sterol biosynthesis strains

Strain	Maximum doubling time (λ, h)	Doubling time vs WT	Growth rate (μ, h^{-1})	Growth rate vs WT
BY4741	1.415	1.00	0.239	1.00
HsS1.1	1.315	0.93	0.214	0.90
HsS1.2	4.502	3.18	0.047	0.20
HsS1.3	3.866	2.73	0.044	0.18
HsS1.4	1.577	1.11	0.176	0.74
HsS1.5	2.720	1.92	0.135	0.56
HsS1.6	4.352	3.08	0.114	0.48
HsS1.7	3.093	2.19	0.135	0.56
HsS1.8	5.123	3.62	0.084	0.35
HsS5.1	3.278	2.32	0.051	0.21
HsS6.1	2.710	1.92	0.040	0.17
HsS7.1	3.538	2.50	0.007	0.03

Strain	Maximum doubling time (λ, h)	Doubling time vs WT	Growth rate (μ, h^{-1})	Growth rate vs WT
HsS8.1	1.766	1.25	0.063	0.26
HsS9.1	6.104	4.31	0.056	0.23
HsS9.2	6.622	4.68	0.021	0.09

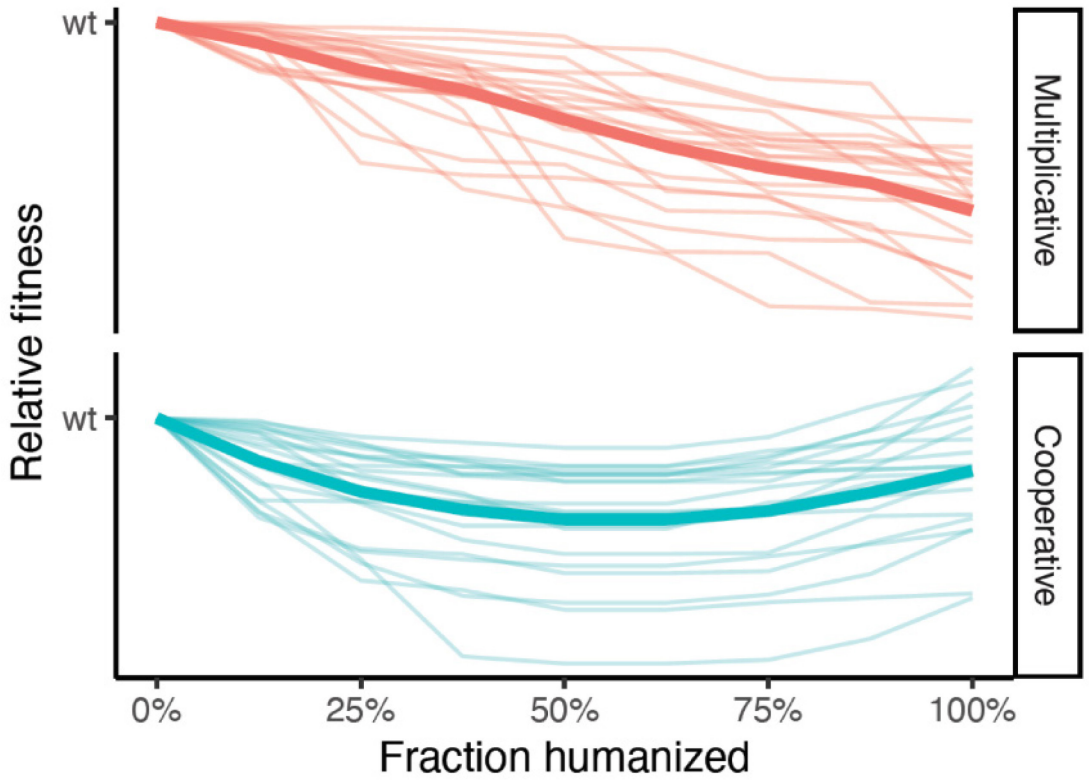


Figure 13: Illustrative simulation of the multiplicative and cooperative models for higher order humanization

Each thin line shows the fitness over the hypothetical course of a whole-pathway humanization project for a module with 8 genes. Thick lines represent averages. In the multiplicative model, fitness of each gene is independent of the other humanizations, and final fitness is given by the product of all individual fitness alterations. In the cooperative model, fitness is dependent on the fraction of humanized genes in that module, resulting in more complex fitness losses and gains. *From Azat Akhmetov*[134].

While growth curves can quantify the overall fitness trends, spotting assays using serial dilutions allow for qualitative observation of growth phenotypes at the single-colony level. As with the growth curves, the single-gene humanized strains showed distinct phenotypic variations in response to temperature stress. HsS1.2, HsS1.3, and HsS1.8 in particular were very susceptible to the low temperature growth condition (Figure 14). For instance, the yeast strains HsS1.2 and HsS1.3, and HsS1.6 and HsS1.7 to a lesser degree, grew at a slower rate than the wild-type colonies. For the multi-gene humanized strains, HsS5.1 was remarkably tolerant to temperature stress in either direction, though the 25°C condition did slow the growth of single colonies significantly. Interestingly, HsS7.1 and HsS8.1 behaved similarly in response to temperature stress and had similar growth rates of single colonies, despite their differing growth rates in liquid culture. HsS9.1 and HsS9.2 were both extremely sensitive to temperature change in either direction, showing almost no growth. In addition to reduced stress resistance in these strains, disruption of sterol biosynthesis can lead to increased weak acid sensitivity in strains, which is pertinent due to the lithium acetate involved in making transformation-competent cells, and may further explain the difficulty of humanizing additional genes in these strains[135–137].

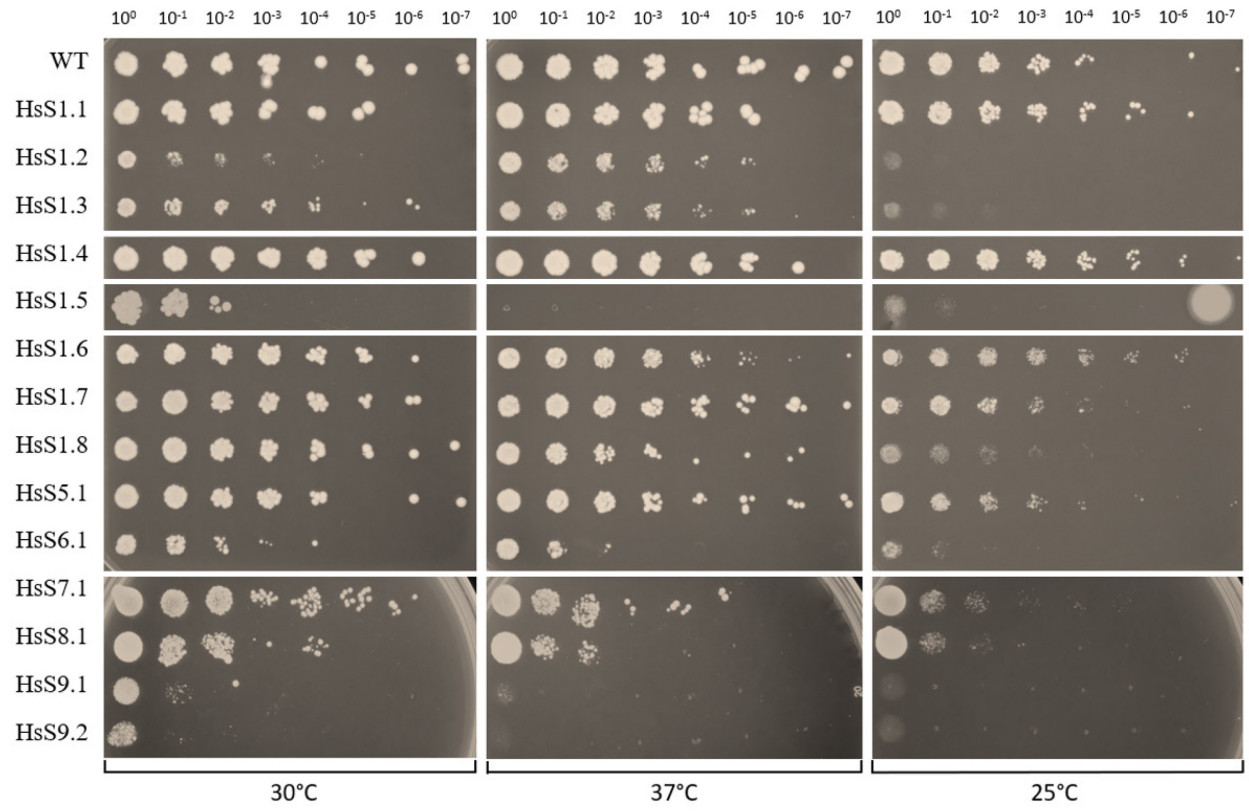


Figure 14: Spotting assays of humanized strains

All HsS strains used or built in this study were assessed under three different temperature conditions for growth efficiency at different seeding concentrations (OD_{600}, listed at the top). 30°C is the normal growth condition for yeast – reduced growth at higher or lower temperatures compared to the wild-type indicates strains have a low-tolerance to environmental stress, which is a common phenotype when there is dysfunction of the ergosterol biosynthesis pathway.

4.3 Differential protein expression in humanized strains

While PCR and sequencing validate the correct integration of the humanized genes, it does not reveal whether these genes show optimal expression. Fitness defects observed in humanized strains could partly be attributed to lower/unstable human protein expression compared to their corresponding yeast equivalents. Of particular interest was the significant growth rate recovery from strain HsS7.1 towards strain HsS8.1, involving the humanization of *HsFDFT1*, the human gene replaceable at the native locus only in a partially-humanized sterol strain.

In an effort to both confirm protein expression of the human genes and investigate proteome-wide changes induced by these successive rounds of humanizations, liquid chromatography-mass spectrometry (LC/MS) was performed on whole-cell lysates of the wild-type, HsS7.1 and HsS8.1 strains. LC/MS identified an average of 1445 distinct yeast proteins per lysate sample, and a total of 6062 unique proteins across all samples. Of the sixteen sterol biosynthesis genes being evaluated in this study, all were detected in at least one of the replicates for the wild-type lysates with the exception of ScERG7 and ScHMG2 (Figure 15). In the humanized strains HsS7.1 and HsS8.1, again with the exception of ScERG7 and ScHMG2, all yeast genes that had been replaced with human genes were absent in replicates, and the corresponding human proteins were present. Expression of a number of sterol pathway yeast proteins (ScERG10, ScMVD1, ScERG20, and ScERG27) increased in the humanized strains compared to the wild-type. ScERG9 also showed increased expression in HsS7.1, before being replaced with the human gene in strain HsS8.1 (Figure 16).

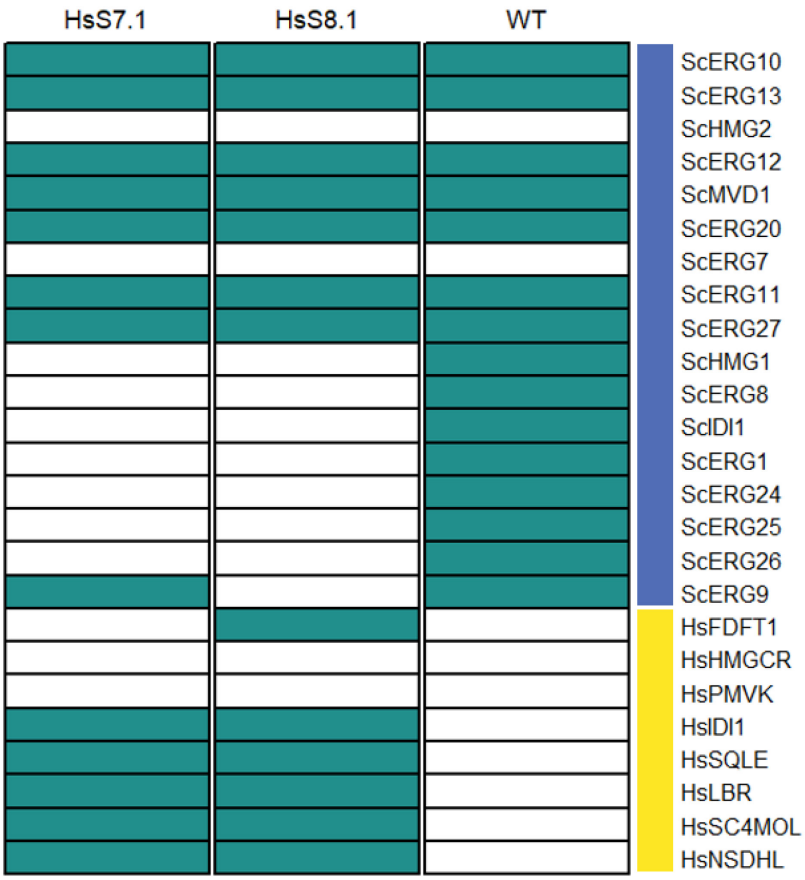


Figure 15: Detection of native yeast and replaced human proteins in humanized and wild-type strains

Whole-cell protein lysates were analyzed by LC/MS, and unique spectrum counts were used to identify human and yeast peptide sequences in two engineered strains (HsS7.1 and HsS8.1) and a wild-type strain (WT). All yeast genes in the sterol biosynthesis pathway but two (ScHMG2, ScERG7) are detected in the wild-type; in HsS7.1 and HsS8.1, yeast proteins are absent where the native gene has been replaced with a human ortholog, and the human genes are detected instead with the exception of two (HsHMGCR and HsPMVK). Coloured boxes indicate detection of protein in at least one of the three replicate samples, white boxes indicate no protein detection in any of the replicates. Blue bar indicates yeast genes, and yellow indicates human genes.

To detect human proteins, peptide reads were also mapped against human proteome database. Five of seven human proteins were identified in HsS7.1, and six of eight were identified in HsS8.1 – in both strains, HsHMGCR and HsPMVK were not detected (Figure 15). HsFDFT1, the previously difficult-to-replace gene introduced into strain HsS8.1, was only detected in one of the three replicates, compared to multiple spectra of ScERG9 detected in the wild-type and HsS7.1 strains, suggesting that even the low expression of that human gene was able to function.

Expression of HsIDI1 remained high relative to other yeast proteins in the pathway, as did ScIDI1 in the wild-type strain. Expression of HsLBR and HsSQLE remained relatively low and consistent in both strains, while expression of HsSC4MOL and HsNSDHL decreased in HsS8.1 to levels comparable with the low expression of ScERG25 and ScERG26 in the wild-type strain, indicating these proteins were up-regulated in HsS7.1, which may partly explain the growth defect. While further study will be required to confirm whether HsHMGCR is being expressed at a very low level or not at all, or if that step of the pathway is being catalyzed entirely by ScHMG2, its lack of detectability in this data provides a potential clue for the low growth rates of the humanized strains. Given that HMG-CoA reductase represents the rate-limiting step of ergosterol biosynthesis, reduced protein expression of the human protein could be significantly limiting the rate of metabolite flow in this essential metabolic pathway.

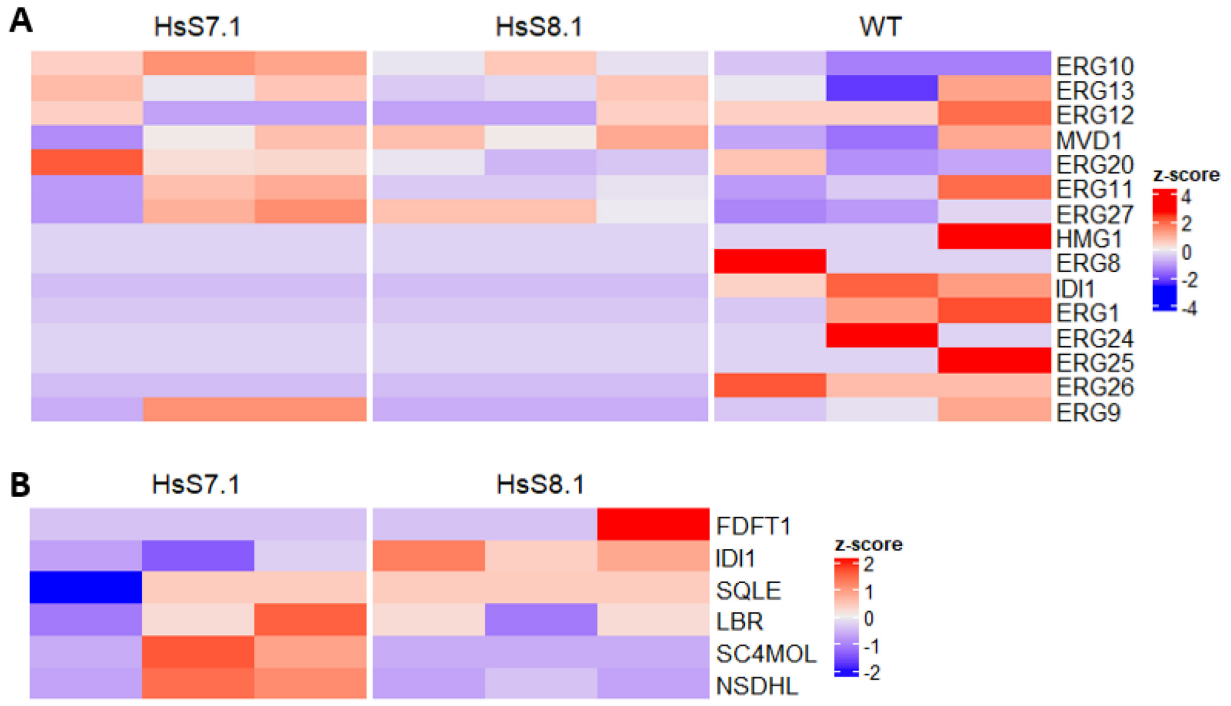


Figure 16: Differential expression of yeast and human sterol biosynthesis proteins in humanized strains

Z-scores were computed based on unique spectrum counts from whole cell protein lysate after LC/MS across groups for (A) yeast sterol biosynthesis proteins and (B) human sterol biosynthesis proteins

expressed in two humanised strains (HsS7.1 and HsS8.1) compared to the wild-type. The colour scale describes relative shifts in protein expression between samples (red = increased, blue = decreased). Each replicate is its own column. Human proteins are only present in the humanised strains, but with expression differences between strains. Yeast genes that have been replaced with the human counterparts are not expressed in HsS7.1 and HsS8.1.

Differential expression of proteins was also assessed at a proteome level as a preliminary step to investigate whether the humanization of the yeast sterol biosynthesis system has cascading effects on other pathways. Fold changes were mapped between strain pairs: HsS7.1 to wild-type, HsS8.1 to wild-type, and HsS8.1 to HsS7.1, to broadly assess the up- or down-regulated proteins/pathways. Compared to the wild-type, HsS7.1 showed 48 significantly upregulated and 57 significantly downregulated genes (Figure 17A). HsS8.1 showed more alterations in protein expression, with 111 significantly upregulated and 101 significantly downregulated genes (Figure 17B). From HsS7.1 to HsS8.1, 55 proteins were significantly upregulated and 56 downregulated, with fewer overall high-significance fold-changes than when strains were compared to the wild-type (Figure 17C). Further investigations will be needed to determine specific mechanisms, but the significant increase in expression of proteins such as GCY1, AMS1, and MCR1, may suggest there are yet-to-be-discovered interactions between these genes and the sterol biosynthesis pathway. As one example, several of these genes (GCY1, GDH2, TDH1) are oxidoreductases involved in carbon metabolism – these enzymes may be upregulated in an attempt in response to stress from reduced ergosterol[128]. Sequential humanization in the sterol pathway may directly alter protein expression levels by as-yet unknown interactions or regulation, or indirectly by changing sterol composition in the cells, leading to pleiotropic changes.

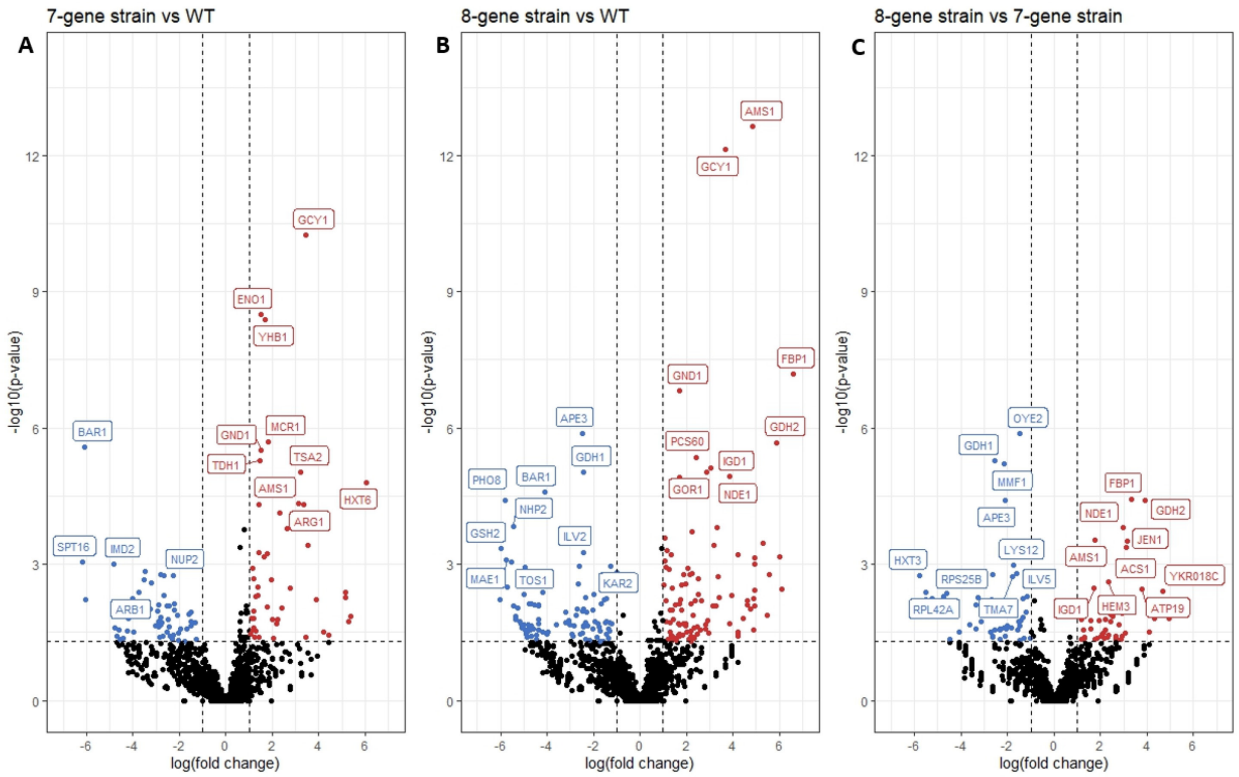


Figure 17: Altered individual protein expression in humanized strains

Using unique spectrum counts after LC/MS of two humanised strains (HsS7.1 and HsS8.1) and a wild-type strain, the edgeR differential expression package computed protein expression fold changes between each pair of strains: (A) HsS7.1 compared to the wild-type; (B) HsS8.1 compared to the wild-type; (C) HsS8.1 compared to HsS7.1. Coloured data points represent genes with a fold-change greater than 1 (red) or less than -1 (blue), with a cut-off of $p < 0.05$. Names in boxes represent a selection of the genes with the most significant fold changes and highest amplitude of change.

Enriched gene pathways, from the Kyoto Encyclopedia of Genes and Genomes, were also mapped to genes with significant fold-changes and p-values. In all strain comparisons, the biosynthesis of secondary metabolites pathways, including amino acids, cofactors, and 2-oxocarboxylic acids, are enriched among the altered protein expressions detected (Figure 18). Carbon metabolism, including glyoxylate and dicarboxylate metabolism and glycolysis, is enriched in both humanized strains when compared to the wild-type, as well as a number of unique amino acid and lipid metabolism pathways. Comparing the two humanized strains, from HsS7.1 to HsS8.1 with significant fitness improvement, oxidative phosphorylation is the most enriched pathway. This indicates that replacing *ERG9* with *FDFT1* in strain HsS8.1 has a positive impact

on the energy metabolism in this strain and may explain the fitness improvement. It is not clear whether the proteome-wide changes in these pathways are due to structural factors, such as reduced or altered ergosterol synthesis, or altered regulation, such as some sterol biosynthesis proteins potentially having key, unknown regulatory interactions with genes and proteins in these pathways. Further study will elucidate the metabolic landscape and reasons for the fitness costs and pathway mis-regulation in these strains. Overall, these results show that sequential humanization of pathways will reveal novel interactions between proteins, not just at the pathway level, but also at the proteome level.

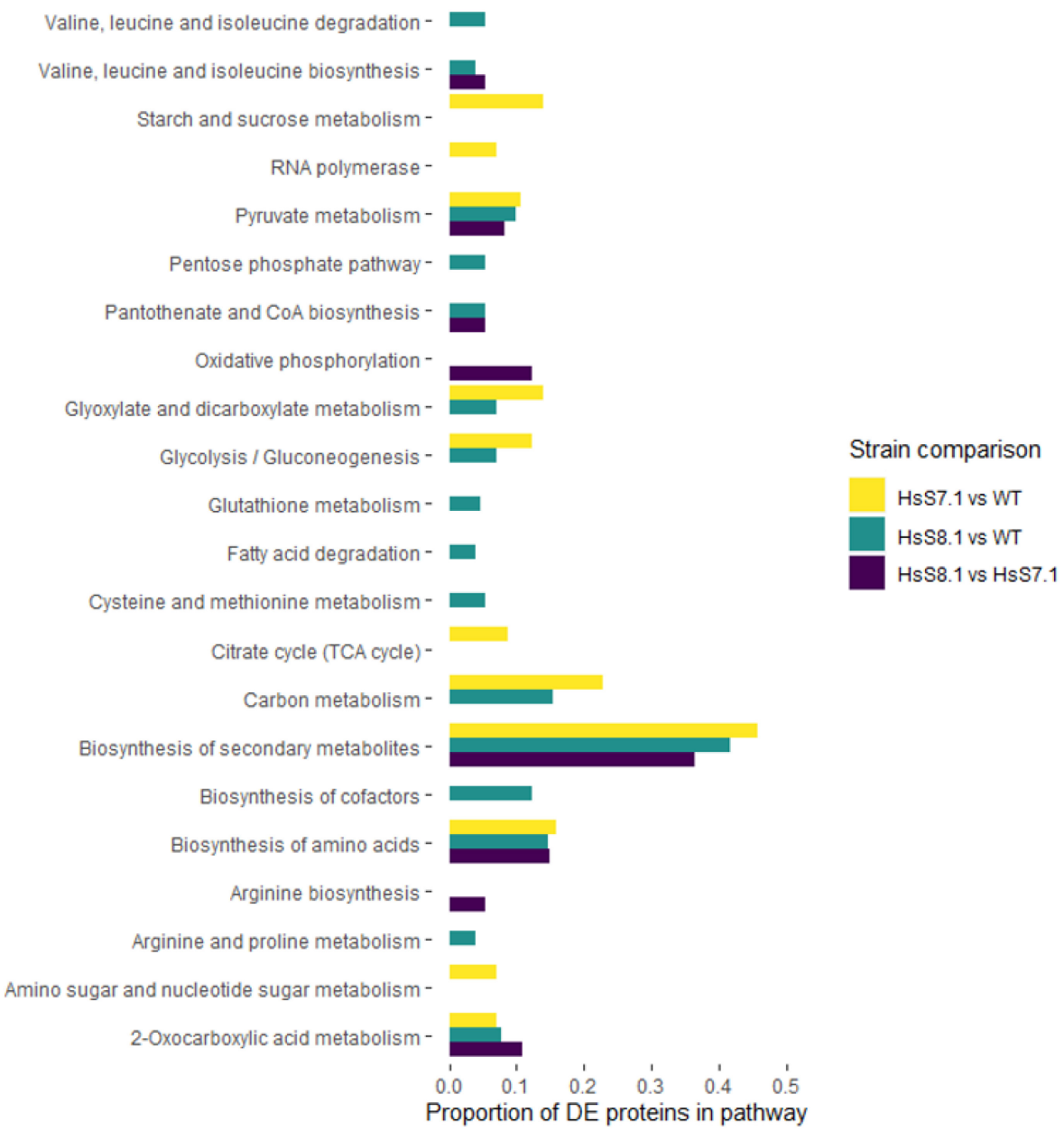


Figure 18: KEGG pathway enrichment in altered protein expression in humanized strains

Proteins with significant absolute fold-changes > 1 (differentially expressed (DE) proteins) were mapped to KEGG pathway associations to select the most enriched pathways in comparisons between HsS7.1 and wild-type (yellow), HsS8.1 and wild-type (green), and HsS8.1 and HsS7.1 (purple). The fraction of the total DE proteins mapped to each pathway is mapped on the x-axis.

4.4 Differential localization of human and yeast orthologs expressed in yeast

While the subcellular localization of certain human sterol biosynthesis proteins is unclear, the GFP-tagged protein studies have elucidated the localization of most of these proteins in yeast and humans. Sterol proteins, in either organism, are localized to the cytoplasm or the endoplasmic reticulum membrane, particularly the proteins involved in the downstream pathway, where the sterol precursors are highly hydrophobic and membrane-bound. Particularly in humans, many sterol biosynthesis proteins are sorted into lipid particles when overexpressed or when sterol concentrations are high, as a feedback mechanism to control sterol levels in the cell[138]. Due to these precise localization and expression contexts, we hypothesized that subcellular localization of proteins could be an important component informing replaceability of humanized genes. For instance, HsACAT1, the first protein of the cholesterol biosynthesis pathway, contains a mitochondrial localization tag and is frequently detected in the mitochondrial membrane (Table 5). Comparatively, the orthologous yeast protein ScERG10 is predominantly cytoplasmic. Therefore, if the human gene is not expressed in the same organelle in yeast, it may contribute to the lack of replaceability in yeast when expressed from the native promoter. To investigate whether the yeast and human orthologs of non-replaceable genes are differentially localized in yeast, I used the yeast GFP expression collection to confirm subcellular localization of the native proteins in yeast[9]. I also created GFP-tagged human gene expression vectors under the control of a constitutive yeast GPD promoter, to investigate localization of human proteins in yeast.

Table 6: Localization of GFP-tagged yeast and human proteins in their respective cell types

Yeast protein	Human protein	Yeast localization[1]	Human localization[2]	Replaceability at native locus
ERG10	ACAT1	Nucleus, cytoplasm	Mitochondria	No
ERG13	HMGCS1	Nucleus, cytoplasm	Nucleus, cytoplasm	No
HMG1	HMGCR	Nucleus, cytoplasm	ER, peroxisome	Yes[3]

Yeast protein	Human protein	Yeast localization[1]	Human localization[2]	Replaceability at native locus
ERG12	MVK	Cytoplasm	Cytoplasm	Yes
ERG8	PMVK	Cytoplasm	Cytoplasm	Yes
MVD1	MVD	Cytoplasm	Cytoplasm	No
IDI1	IDI1	Cytoplasm	Cytoplasm, peroxisome	Yes
ERG20	FDPS	Cytoplasm	Cytoplasm	No
ERG9	FDFT1	ER	ER	Yes
ERG1	SQLE	ER	ER	Yes
ERG7	LSS	ER	ER	No
ERG11	CYP51A1	ER	ER	Yes
ERG24	LBR	ER	ER	Yes
ERG25	SC4MOL	ER	ER	Yes
ERG26	NSDHL	ER	ER	Yes
ERG27	HSD17B7	ER	ER	No

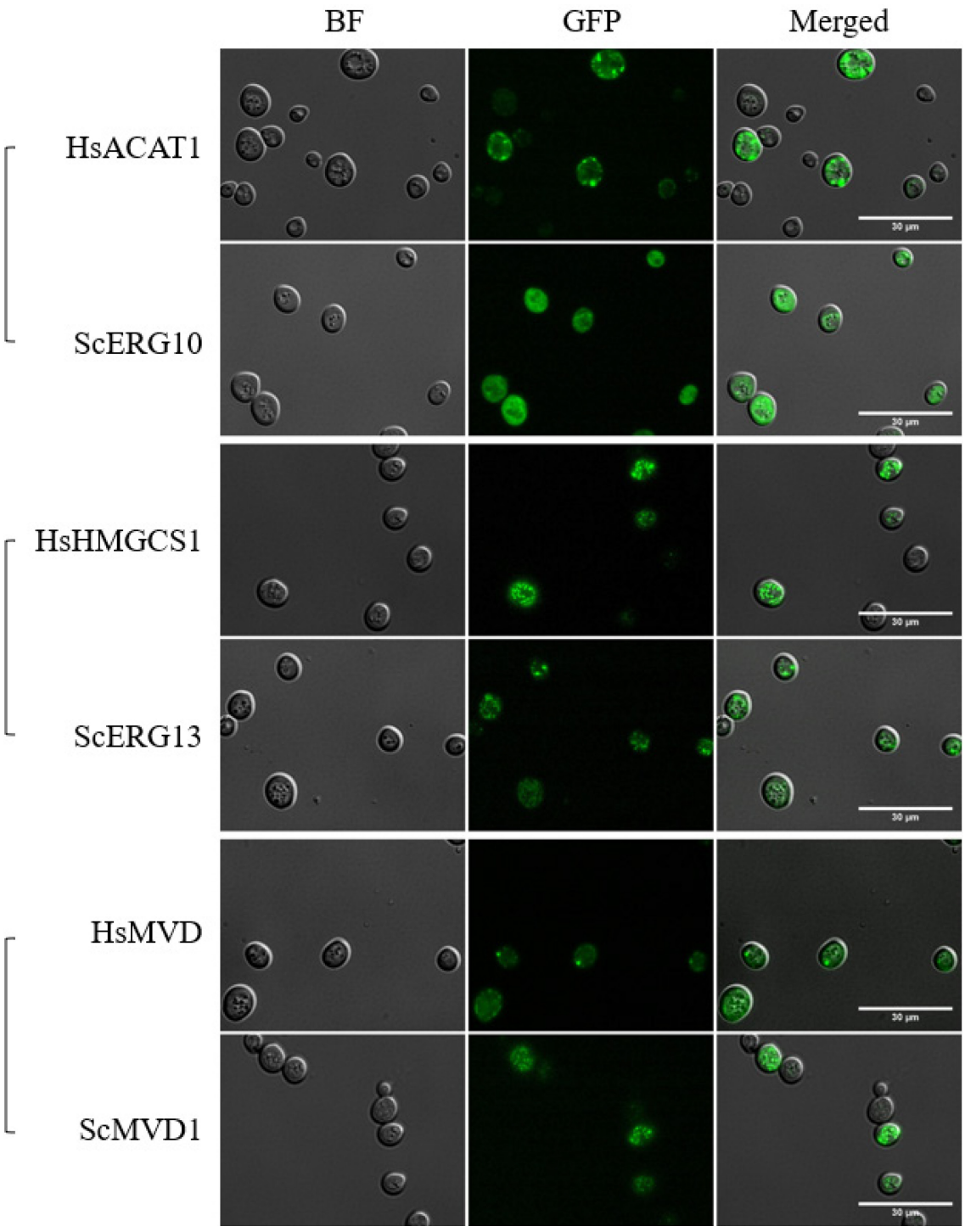
1 Yeast localization data obtained from the CYCLoPs database on thecellvision.org, or from the Saccharomyces Genome Database (SGD) if not available

2 Human localization data in human cells obtained from uniport.org

3 HMGCR replaces HMG1 at the native locus, but the paralog HMG2 is still functional, and has not yet been successfully replaced in conjunction with HMG1

GFP-fluorescence for all six human proteins indicates that while replaceability has yet to be determined, yeast do tolerate expression of each of these human genes (i.e., they are not toxic to yeast, or degraded quickly by the cell, and should be functional) (Figure 19). The ortholog pairs HsACAT1/ScERG10, HsHMGCS1/ScERG13, HsMVD/ScMVD1, and HsLSS/ScERG7 all show clear differential localization in yeast (Figure 19). For instance, ScERG10 is expressed throughout the cytoplasm, whereas the HsACAT1 is aggregated in smaller structures – these could be mitochondria, or lipid particles. While the precise subcellular locations are not clear at this time, the ortholog pairs HsHMGCS1/ScERG13, HsMVD/ScMVD1, and HsLSS/ScERG7 are all differentially localized in the cell to some degree, supporting the hypothesis that differential localization may be a bottleneck when humanizing at the native locus, particularly when protein

expression is highly regulated. One limitation of expressing proteins from a constitutive strong promoter is that it doesn't match the native context, and excess proteins may be localized into lipid droplets or form non-functional aggregates[138–140]. In the future, these microscopy experiments will be repeated at different phases of cell growth, coupled with localization tags to clarify which structures are being observed (e.g., to differentiate mitochondria from the ER). Interestingly, the ortholog pair HsHSD17B7/ScERG27 is not differentially localized in this assay – both proteins distribute throughout the cytoplasm, suggesting localization is not the only bottleneck in humanization. Several ortholog pairs are replaceable when using plasmid-based complementation assays (HsMVD/ScMVD, HsLSS/ScERG7, and HsHSD17B7/ScERG27), indicating that expression may play a role in the non-replaceability at the native locus.


BF
GFP
Merged
HsACAT1
ScERG10
HsHMGCS1
ScERG13
HsMVD
ScMVD1
30 µm

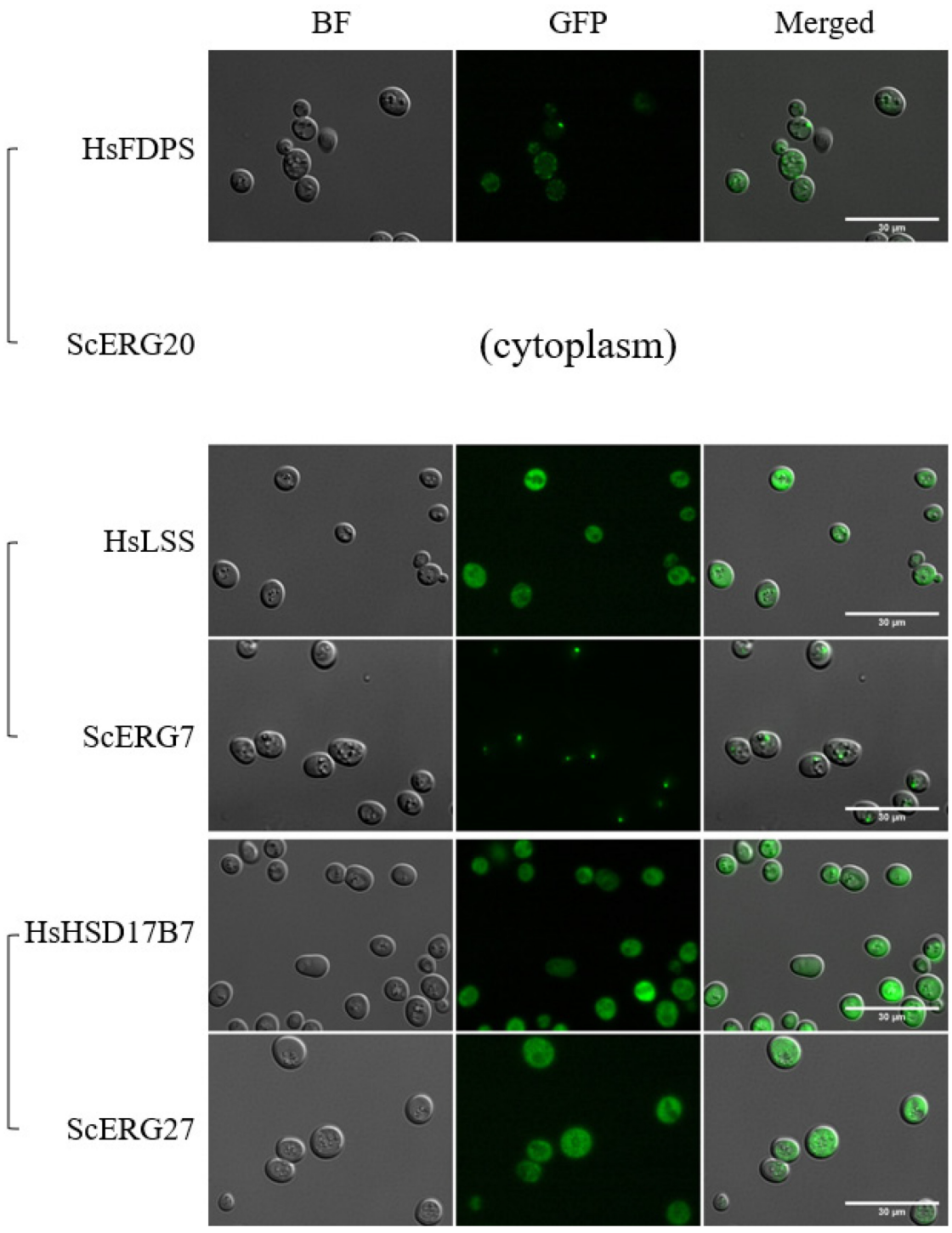


Figure 19: Differential localization of human and yeast orthologous proteins in *S. cerevisiae*

Wild-type yeast were transformed with plasmids expressing GFP-tagged human genes (Hs) and grown to stationary phase. For comparison, yeast from the GFP-tagged collection were grown to the same OD600 [9]. Yeast was washed with PBS and imaged at 630x total magnification. Images represent bright-field, GFP-activation, and overlap. Scale bar = 30µm. Square brackets indicate ortholog pairs for comparison. BF = bright field, GFP = GFP fluorescence.

Chapter 5 – Conclusions and Future Directions

5.1 Future directions

At the moment of the writing of this thesis, I have humanized nine sterol biosynthesis pathway genes in two different combinations in a single yeast strain, one with a functionally complementing *CYP51A1* and the other with a functionally complementing *MVK* in addition to eight other genes. Efforts to humanize more combinations towards building a fully humanized sterol biosynthesis pathway in yeast are underway. Towards humanizing the remaining six genes that have yet to be replaced at the native locus, we will employ the ectopic chromosomal repair template strategy described above, as well as additional rounds of sequential humanization. Alternatively, we will proceed with plasmid-based complementation for the remaining six genes by expressing all six as a module in yeast.

While the mass spectrometry confirmed protein expression of almost all of the sterol biosynthesis genes, coverage was relatively low, and could be improved by better cell lysate processing methods and longer run times in the future. For specific genes of interest, biochemical assays such as Western blots and antibody tagging could be used to further characterize expression among humanized strains. We performed mass spectrometry analysis on humanized HsS7.1 and HsS8.1 strains only, between which there was a significant fitness recovery. Mass spectrometry of additional strains as they are generated will help to systematically discover altered protein expression, potentially unlocking novel genetic interactions, among pathway members and other genes in the cell. Not only will these inform further how yeast behave when sterol biosynthesis is mis-regulated/humanized, but also unveil possible bottlenecks to humanization of entire pathways in yeast, identifying distinctly regulated sterol genes in humans and yeast. While mass spectrometry was chosen specifically to confirm expression of functional proteins in these strains,

in the future RNA-sequencing will be investigated to further elucidate global gene expression changes in humanized strains.

One of the major challenges of humanizing the entire sterol biosynthesis pathway is that most of the genes in it are essential, and replacing these genes with conserved human versions which have evolved in a species-specific manner can exact a high fitness toll. For example, one of the major reasons for the non-replaceability of *HsFDPS* could be that its specific catalytic function is essential not only to sterol biosynthesis, but also to the isoprenoid biosynthesis pathway. If the function of this human gene is too different from that of its yeast ortholog, multiple essential metabolic pathways are impacted, which may be too high a cost for the yeast to bear. The individually non-replaceable genes downstream of isoprenoid biosynthesis – *HsFDFT1/ScERG9, HsLSS/ScERG7, HsHSD17B7/ScERG27* – may also be exacting too high a fitness cost when expressed from native promoters rather that by strong heterologous promoters. However, the later section of the pathway (from *ScERG9* onwards) can be made non-essential by supplementing ergosterol to the medium, and growing yeast in anaerobic conditions. We hypothesize that this strategy may partially relieve the fitness cost, and allow these genes to be humanized at the native locus. These strains could then be further tested on media lacking ergosterol for functional replaceability. This strategy also eliminates two fitness challenges towards humanization using CRISPR, i.e. – maintaining sterol biosynthesis, and repairing DSBs at essential loci.

As for the upstream non-replaceable genes – *HsACAT1/ScERG10, HsHMGCS1/ScERG13, HsFDPS/ScERG20* – supplementing intermediate metabolites could also relieve the fitness cost of altered sterol biosynthesis regulation. However, they will remain essential for isoprenoid biosynthesis. In humans, *HsACAT1* catalyzes the first step of cholesterol biosynthesis, and is highly regulated by cellular concentrations of cholesterol. Recently, a yeast strain that converts

zymosterol to cholesterol, was engineered using zebrafish orthologous genes[84]. Attempting to humanize *HsACAT1/ScERG10* in a cholesterol-producing strain could inform whether this key regulatory interaction is impeding replaceability at the native locus.

Taken together, the experiments performed here to engineer and characterize novel humanized strains in the sterol biosynthesis pathway in yeast start to reveal a picture that highlights essential regulatory elements of the pathway (Figure 20). While most of the genes in this pathway are replaceable by plasmid overexpression of the human ortholog, replacements at the native locus represent an additional challenge. The regulation of *ScERG10* expression by ergosterol, much like *HsACAT1* is regulated by cholesterol levels in human cells, suggests that regulation of this step is critical for cell survival. While *HMG1,* a highly regulated step in sterol biosynthesis, is efficiently replaced by *HsHMGCR*, the duplicated yeast *HMG2* copy must be playing a role in maintaining strain fitness. Further investigation as to whether a dually replaced strain, where a copy of *HMGCR* is also inserted at the *HMG2* locus, could reveal the extent to which regulation of this step is essential. The consistent growth-delays observed in strains with a humanized *MVK* suggest a severe fitness cost associated with the altered function or regulation of this gene. *ERG7* and *ERG27* are known to interact, and the difficulty of replacing either one of them under native locus control may suggest this interaction is essential and a bottleneck in humanization assays[141]. And finally, the inability of *ERG20* to be replaced under any known conditions suggests this to be a regulatory keystone of sterol biosynthesis, as well as heme and ubiquinone synthesis, and prenylated proteins[78–80]. While it remains to be seen whether this gene becomes replaceable as we progressively humanize the sterol biosynthesis pathway in yeast, additional study may reveal novel regulatory roles of this gene in humans.

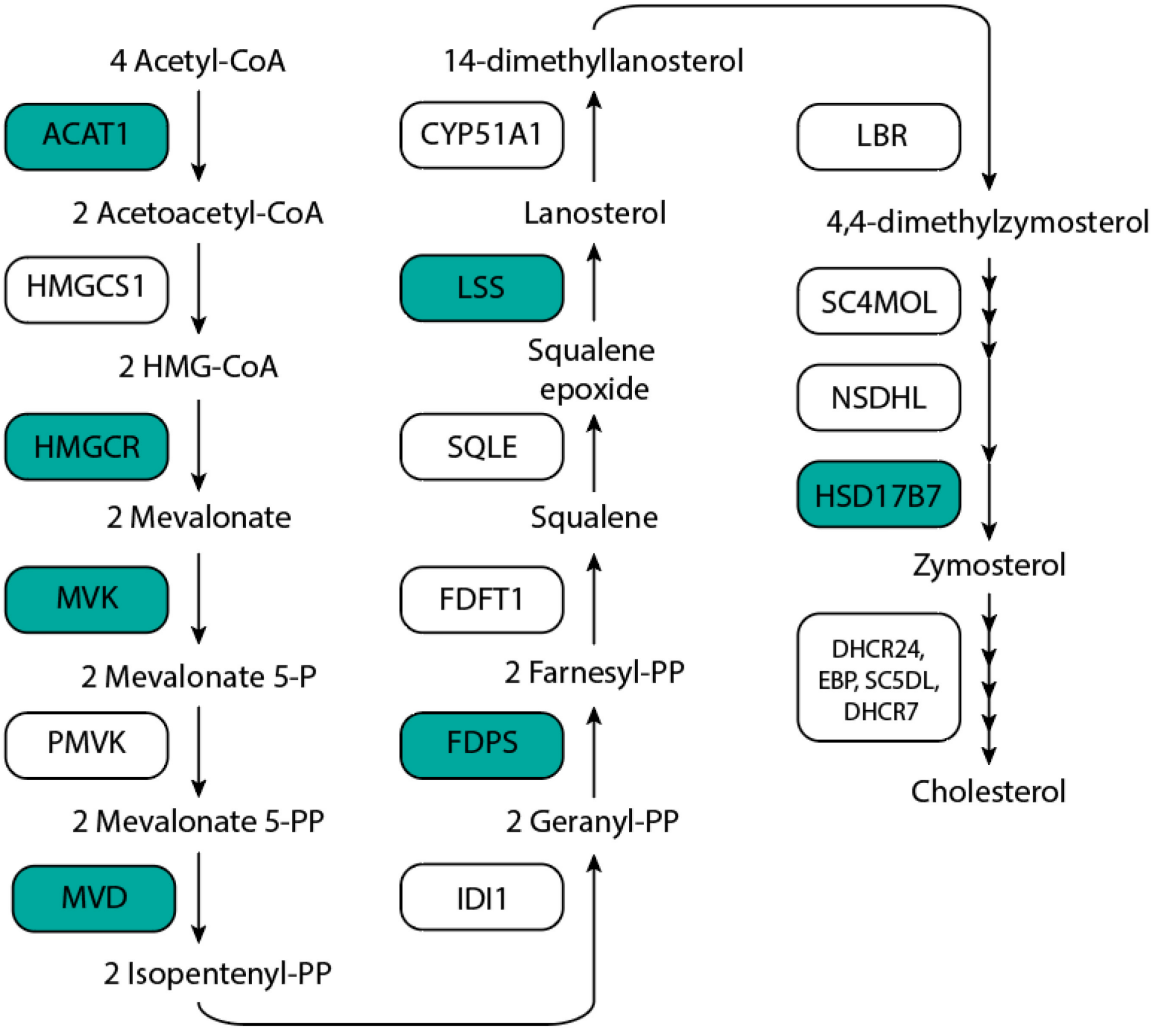


Figure 20: Humanization of the sterol biosynthesis pathway reveals bottlenecks towards whole-pathway engineering

Cholesterol biosynthesis in humans a multi-step pathway comprising 20 enzymes – of the upstream pathway, all 16 genes have yeast orthologs, of which we know 10 are replaceable at the native locus to date, and all but FDPS are replaceable using plasmid-based complementation. Sequential humanization of the sterol biosynthesis pathway in yeast has revealed a number of enzymes as significant regulatory nodes (highlighted in blue), which are difficult to humanize, possibly due to mislocalization or expression, or present severe fitness defects in multiple strains.

5.2 Conclusion

In this study, we undertook the challenge of humanizing the entire sterol biosynthesis pathway in *Saccharomyces cerevisiae*, building upon past efforts by collaborators at the University of Texas at Austin. Due to the many human genetic disease associations, a need for platforms to discover novel drug targets, and a high functional conservation of this module between yeast and humans, the rationale for engineering the entire sterol biosynthesis pathway is evident. In this study, I individually and sequentially humanized multiple genes in the sterol biosynthesis pathway in yeast using CRISPR-Cas9, demonstrating the utility of this approach for engineering pathways in yeast. These efforts resulted in the generation of four novel individually-humanized strains, and two novel strains with nine humanized genes each, as well as intermediate strains with different combinations of pathway members. The humanization of *HsFDFT1* only after several other human genes were integrated demonstrates requirement of local interactions for functional replaceability of certain genes.

I investigated the comparative fitness of these strains using growth curves to observe where potential bottlenecks in the pathway may be occurring, such as in the severe growth lag seen when *HsMVK/ScERG12* is replaced, both individually and in a multi-gene strain. I used mass spectrometry to confirm expression of the yeast and human proteins in these strains, and provided initial evidence that humanizing an entire pathway has proteome-wide consequences in the cell, which may reveal novel genetic interactions between sterol biosynthesis and other pathways. I elucidated differential expression of human and yeast ortholog pairs under constitutive expression, highlighting altered localization of proteins as another potential bottleneck in pathway engineering efforts that needs further validation. Finally, I developed a new strategy for addressing genes which are difficult to engineer or replace in yeast.

This study is just one among many showing that yeast remains an exceptionally powerful organism for study of human genetics, basic biology, and evolution in a simplified context. The strains engineered and characterized in this study will be useful as platforms to study gene-gene or gene-phenotype associations in human genetic diseases, and discover novel drugs to treat genetic diseases in these pathways. While humanization of the entire pathway has not yet been achieved, major steps towards a fully humanized essential sterol biosynthesis pathway in yeast have been made. Furthermore, even these intermediate steps highlight the wealth of information that efforts to humanize whole pathways can provide to synthetic biologists working towards engineering complex pathways in yeast.

Chapter 6 – Bibliography

1. Kachroo, A. H., Vandeloo, M., Greco, B. M. & Abdullah, M. Humanized yeast to model human biology, disease and evolution. *Dis Model Mech* **15**, (2022).
2. Goffeau, A. *et al.* Life with 6000 genes. *Science (1979)* **274**, 5287 (1996).
3. Fraczek, M. G., Naseeb, S. & Delneri, D. History of genome editing in yeast. *Yeast* **35**, 361 (2018).
4. Aylon, Y. & Kupiec, M. New insights into the mechanism of homologous recombination in yeast. *Mutat Res Rev Mutat Res* **566**, 231–248 (2004).
5. Nielsen, J. Yeast Systems Biology: Model Organism and Cell Factory. *REVIEW* (2019) doi:10.1002/biot.201800421.
6. Balakrishnan, R. *et al.* YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database (Oxford)* **2012**, (2012).
7. Dwight, S. S. *et al.* Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res* **30**, 69–72 (2002).
8. Giaever, G. & Nislow, C. The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics* **197**, 451–465 (2014).
9. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature 2003 425:6959* **425**, 686–691 (2003).
10. Cuperus, J. T., Lo, R. S., Shumaker, L., Proctor, J. & Fields, S. A tetO toolkit to alter expression of genes in Saccharomyces cerevisiae. *ACS Synth Biol* **4**, 842 (2015).
11. Leeuwen, J. van *et al.* Systematic analysis of bypass suppression of essential genes. *Mol Syst Biol* **16**, e9828 (2020).
12. Arita, Y. *et al.* A genome-scale yeast library with inducible expression of individual genes. *bioRxiv* 2020.12.30.424776 (2021) doi:10.1101/2020.12.30.424776.
13. Lee, M. E., Deloache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. (2015) doi:10.1021/sb500366v.
14. Kuzmin, E. *et al.* Synthetic Genetic Array Analysis for Global Mapping of Genetic Networks in Yeast. *Methods in Molecular Biology* **1205**, 143–168 (2014).
15. Luo, X. *et al.* Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature 2019 567:7746* **567**, 123–126 (2019).
16. Madhavan, A. *et al.* Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microbial Cell Factories 2021 20:1* **20**, 1–17 (2021).
17. Kavščček, M., Stražar, M., Curk, T., Natter, K. & Petrovič, U. Yeast as a cell factory: current state and perspectives. *Microb Cell Fact* **14**, 94 (2015).
18. Zhao, Y., Zhang, Y., Nielsen, J. & Liu, Z. Production of β-carotene in Saccharomyces cerevisiae through altering yeast lipid metabolism. *Biotechnol Bioeng* **118**, 2043–2052 (2021).
19. da Silva, N. A. & Srikrishnan, S. Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. *FEMS Yeast Res* **12**, 197–214 (2012).
20. Yen-Ting-Liu *et al.* The partitioning and copy number control systems of the selfish yeast plasmid: an optimized molecular design for stable persistence in host cells. *Microbiol Spectr* **2**, (2014).
21. Bourgeois, L., Pyne, M. E. & Martin, V. J. J. A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast. *ACS Synth Biol* **7**, 2675–2685 (2018).
22. Niu, W., Guo, J. & van Dien, S. Heterologous pathway engineering. *Metabolic Engineering for Bioprocess Commercialization* 31–52 (2016) doi:10.1007/978-3-319-41966-4_3/FIGURES/2.

23. Laurent, J. M., Young, J. H., Kachroo, A. H. & Marcotte, E. M. Efforts to make and apply humanized yeast. *Brief Funct Genomics* **15**, 155–163 (2016).
24. Kachroo, A. H. *et al.* Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science (1979)* **348**, 921–925 (2015).
25. Giaever, G. *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature 2002 418:6896* **418**, 387–391 (2002).
26. VanderSluis, B. *et al.* Integrating genetic and protein-protein interaction networks maps a functional wiring diagram of a cell. *Curr Opin Microbiol* **45**, 170 (2018).
27. McGary, K. L., Lee, I. & Marcotte, E. M. Broad network-based predictability of Saccharomyces cerevisiae gene loss-of-function phenotypes. *Genome Biol* **8**, R258 (2007).
28. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science (1979)* **320**, 362–365 (2008).
29. Costanzo, M. *et al.* A global interaction network maps a wiring diagram of cellular function. *Science* **353**, (2016).
30. Ohnuki, S., Oka, S., Nogami, S. & Ohya, Y. High-Content, Image-Based Screening for Drug Targets in Yeast. *PLoS One* **5**, (2010).
31. Armour, C. D. & Lum, P. Y. From drug to protein: using yeast genetics for high-throughput target discovery. *Curr Opin Chem Biol* **9**, 20–24 (2005).
32. McGary, K. L. *et al.* Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proceedings of the National Academy of Sciences* **107**, 6544–6549 (2010).
33. Guaragnella, N. *et al.* The expanding role of yeast in cancer research and diagnosis: insights into the function of the oncosuppressors p53 and BRCA1/2. *FEMS Yeast Res* **14**, 2–16 (2014).
34. Denoth Lippuner, A., Julou, T. & Barral, Y. Budding yeast as a model organism to study the effects of age. *FEMS Microbiol Rev* **38**, 300–325 (2014).
35. Lasserre, J. P. *et al.* Yeast as a system for modeling mitochondrial disease mechanisms and discovering therapies. *Dis Model Mech* **8**, 509–526 (2015).
36. Zhu, F., Nair, R. R., Fisher, E. M. C. & Cunningham, T. J. Humanising the mouse genome piece by piece. *Nature Communications 2019 10:1* **10**, 1–13 (2019).
37. Dutrow, E. v. *et al.* Modeling uniquely human gene regulatory function via targeted humanization of the mouse genome. *Nature Communications 2022 13:1* **13**, 1–15 (2022).
38. Garge, R. K., Laurent, J. M., Kachroo, A. H. & Marcotte, E. M. Systematic humanization of the yeast cytoskeleton discerns functionally replaceable from divergent human genes. *Genetics* **215**, 1153–1169 (2020).
39. Kataoka, T. *et al.* Functional homology of mammalian and yeast RAS genes. *Cell* **40**, 19–26 (1985).
40. Laurent, J. M. *et al.* Humanization of yeast genes with multiple human orthologs reveals functional divergence between paralogs. *PLoS Biol* **18**, (2020).
41. Sun, S. *et al.* An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome Res* **26**, 670–680 (2016).
42. Kachroo, A. H. *et al.* Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. *Elife* **6**, (2017).
43. Hamza, A., Driessen, M. R. M., Tammpere, E., O'Neil, N. J. & Hieter, P. Cross-Species Complementation of Nonessential Yeast Genes Establishes Platforms for Testing Inhibitors of Human Proteins. *Genetics* **214**, 735–747 (2020).

44. Hamza, A. *et al.* Complementation of yeast genes with human genes as an experimental platform for functional testing of human genetic variants. *Genetics* **201**, 1263–1274 (2015).
45. Yang, F. *et al.* Identifying pathogenicity of human variants via paralog-based yeast complementation. *PLoS Genet* **13**, (2017).
46. Heinisch, J. J. & Brandt, R. Signaling pathways and posttranslational modifications of tau in Alzheimer's disease: the humanization of yeast cells. *Microbial Cell* **3**, 135 (2016).
47. Franssens, V. *et al.* The Benefits of Humanized Yeast Models to Study Parkinson's Disease. *Oxid Med Cell Longev* **2013**, (2013).
48. Marini, N. J., Thomas, P. D. & Rine, J. The Use of Orthologous Sequences to Predict the Impact of Amino Acid Substitutions on Protein Function. *PLoS Genet* **6**, e1000968 (2010).
49. Marzo, M. G. *et al.* Molecular basis for dyneinopathies reveals insight into dynein regulation and dysfunction. *Elife* **8**, (2019).
50. Santos, S. M. *et al.* A Humanized Yeast Phenomic Model of Deoxycytidine Kinase to Predict Genetic Buffering of Nucleoside Analog Cytotoxicity. *Genes 2019, Vol. 10, Page 770* **10**, 770 (2019).
51. Friedmann, Y. *et al.* JX401, A p38alpha inhibitor containing a 4-benzylpiperidine motif, identified via a novel screening system in yeast. *Mol Pharmacol* **70**, 1395–1405 (2006).
52. Yellman, C. M. Precise Replacement of Saccharomyces cerevisiae Proteasome Genes with Human Orthologs by an Integrative Targeting Method. *G3: Genes|Genomes|Genetics* **10**, 3189 (2020).
53. Truong, D. M. & Boeke, J. D. Resetting the yeast epigenome with human nucleosomes. *Cell* **171**, 1508 (2017).
54. Kim, S., Juhee, P., Taekyun, K. & Lee, J. The functional study of human proteins using humanized yeast. *Journal of Microbiology* **58**, 343–359 (2020).
55. Agmon, N. *et al.* Phylogenetic debugging of a complete human biosynthetic pathway transplanted into yeast. *Nucleic Acids Res* **48**, 486 (2020).
56. Lee, C. S. K. *et al.* Humanizing the yeast origin recognition complex. *Nat Commun* **12**, (2021).
57. Sultana, S., Abdullah, M., Li, J., Hochstrasser, M. & Kachroo, A. H. Species-specific protein-protein interactions govern the humanization of the 20S proteasome in yeast. *bioRxiv* 2022.06.20.496808 (2022) doi:10.1101/2022.06.20.496808.
58. Boonekamp, F. J. *et al.* Full humanization of the glycolytic pathway in Saccharomyces cerevisiae. *Cell Rep* **39**, 111010 (2022).
59. Wildt, S. & Gerngross, T. U. The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology 2005 3:2* **3**, 119–128 (2005).
60. Khan, A. H., Bayat, H., Rajabibazl, M., Sabri, S. & Rahimpour, A. Humanizing glycosylation pathways in eukaryotic expression systems. *World J Microbiol Biotechnol* **33**, (2017).
61. Li, X. *et al.* Humanization of Yeasts for Glycan-Type End-Products. *Front Microbiol* **13**, (2022).
62. Zafar, F. Humanized Yeast as a Platform to Measure the Functional Impact of Human Genetic Variation. (Concordia University, 2021).
63. Franssens, V. *et al.* The Benefits of Humanized Yeast Models to Study Parkinson's Disease. *Oxid Med Cell Longev* **2013**, (2013).
64. Akhmetov, A. *et al.* Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. *Bio Protoc* **8**, (2018).
65. Labunskyy, V. M. *et al.* The Insertion Green Monster (iGM) Method for Expression of Multiple Exogenous Genes in Yeast. (2014) doi:10.1534/g3.114.010868.

66. Suzuki, Y. *et al.* Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. *Nature Methods 2011 8:2* **8**, 159–164 (2011).
67. Abdullah, M. *et al.* Rapid, scalable, combinatorial genome engineering by Marker-less Enrichment and Recombination of Genetically Engineered loci (MERGE). *bioRxiv* **496490**, (2022).
68. Jordá, T. & Puig, S. Regulation of Ergosterol Biosynthesis in Saccharomyces cerevisiae. *Genes (Basel)* **11**, 1–18 (2020).
69. Suárez, Y. *et al.* Differential effects of ergosterol and cholesterol on Cdk1 activation and SRE-driven transcription. *Eur J Biochem* **269**, 1761–1771 (2002).
70. Liscum, L. & Munn, N. J. Intracellular cholesterol transport.
71. Sokolov, S. S., Trushina, N. I., Severin, F. F. & Knorre, D. A. Ergosterol Turnover in Yeast:An Interplay between Biosynthesis and Transport. *Biochemistry (Moscow)* **84**, 346–357 (2019).
72. Luo, J., Yang, H. & Song, B.-L. Mechanisms and regulation of cholesterol homeostasis. *Nature Reviews* **21**, 227–245 (2020).
73. Cerqueira, N. M. F. S. A. *et al.* Cholesterol Biosynthesis: A Mechanistic Overview. *Biochemistry* **55**, 5483–5506 (2016).
74. Jacquier, N. & Schneiter, R. Mechanisms of sterol uptake and transport in yeast. *Journal of Steroid Biochemistry & Molecular Biology* **129**, 70–78 (2012).
75. Zinser, E., Paltauf, F. & Daum, G. Sterol composition of yeast organelle membranes and subcellular distribution of enzymes involved in sterol metabolism. *J Bacteriol* **175**, 2853 (1993).
76. Dupont, S. *et al.* Ergosterol Biosynthesis: A Fungal Pathway for Life on Land? *Evolution (N Y)* **66**, 2961–2968 (2012).
77. Peña-Diaz, J. *et al.* Mitochondrial Localization of the Mevalonate Pathway Enzyme 3-Hydroxy-3-methyl-glutaryl-CoA Reductase in the Trypanosomatidae. *Mol Biol Cell* **15**, 1356 (2004).
78. Lange, B. M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A* **97**, 13172 (2000).
79. Miziorko, H. M. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch Biochem Biophys* **505**, 131–143 (2011).
80. Kaminska, J. *et al.* The isoprenoid biosynthetic pathway in Saccharomyces cerevisiae is affected in a maf1-1 mutant with altered tRNA synthesis. *FEMS Yeast Res* **2**, 31–37 (2002).
81. Buhaescu, I. & Izzedine, H. Mevalonate pathway: A review of clinical and therapeutical implications. (2007) doi:10.1016/j.clinbiochem.2007.03.016.
82. Park, J., Matralis, A. N., Berghuis, A. M. & Tsantrizos, Y. S. Human isoprenoid synthase enzymes as therapeutic targets. *Front Chem* **2**, (2014).
83. Tansey, T. R. & Shechter, I. Squalene synthase: structure and regulation. *Prog Nucleic Acid Res Mol Biol* **65**, 157–195 (2001).
84. Bean, B. D. M. *et al.* Functional expression of opioid receptors and other human GPCRs in yeast engineered to produce human sterols. *Nat Commun* **13**, (2022).
85. Jordá, T. & Puig, S. Regulation of Ergosterol Biosynthesis in Saccharomyces cerevisiae. *Genes (Basel)* **11**, 1–18 (2020).
86. Gorin, A., Gabitova, L. & Astsaturov, I. Regulation of cholesterol biosynthesis and cancer signaling. *Curr Opin Pharmacol* **12**, 710 (2012).
87. Espenshade, P. J. & Hughes, A. L. Regulation of Sterol Synthesis in Eukaryotes. *http://dx.doi.org.lib-ezproxy.concordia.ca/10.1146/annurev.genet.41.110306.130315* **41**, 401–427 (2007).

88. Dimster-Denk, D. & Rine, J. Transcriptional Regulation of a Sterol-Biosynthetic Enzyme by Sterol Levels in Saccharomyces cerevisiae. *Mol Cell Biol* **16**, 3981–3989 (1996).
89. Chen, H. *et al.* Regulated degradation of HMG CoA reductase requires conformational changes in sterol-sensing domain. *Nature Communications 2022 13:1* **13**, 1–12 (2022).
90. Taramino, S. *et al.* Interactions of oxidosqualene cyclase (Erg7p) with 3-keto reductase (Erg27p) and other enzymes of sterol biosynthesis in yeast. *Biochim Biophys Acta* **1801**, 156 (2010).
91. Mo, C. & Bard, M. A systematic study of yeast sterol biosynthetic protein–protein interactions using the split-ubiquitin system. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1737**, 152–160 (2005).
92. Kodedová, M. & Sychrová, H. Changes in the Sterol Composition of the Plasma Membrane Affect Membrane Potential, Salt Tolerance and the Activity of Multidrug Resistance Pumps in Saccharomyces cerevisiae. *PLoS One* **10**, (2015).
93. Herman, G. E. Disorders of cholesterol biosynthesis: prototypic metabolic malformation syndromes. *Hum Mol Genet* **12**, R75–R88 (2003).
94. Bhattacharya, S., Esquivel, B. D. & White, T. C. Overexpression or Deletion of Ergosterol Biosynthesis Genes Alters Doubling Time, Response to Stress Agents, and Drug Susceptibility in Saccharomyces cerevisiae. *mBio* **9**, (2018).
95. Johnston, E. J., Moses, T., Rosser, S. J. & Emily Johnston, C. J. The wide-ranging phenotypes of ergosterol biosynthesis mutants, and implications for microbial cell factories. (2019) doi:10.1002/yea.3452.
96. Shalini, K., Kumar, N., Drabu, S. & Sharma, P. K. Advances in synthetic approach to and antifungal activity of triazoles. *Beilstein Journal of Organic Chemistry* **7**, 668 (2011).
97. Maxfield, F. R. & Tabas, I. Role of cholesterol and lipid organization in disease. *Nature 2005 438:7068* **438**, 612–621 (2005).
98. Kimbung, S., Lettiero, B., Feldt, M., Bosch, A. & Borgquist, S. High expression of cholesterol biosynthesis genes is associated with resistance to statin treatment and inferior survival in breast cancer. *Oncotarget* **7**, 59640 (2016).
99. Zhang, Z. *et al.* Genomic variations of the mevalonate pathway in porokeratosis. *Elife* **4**, (2015).
100. Grünert, S. C. *et al.* Clinical presentation and outcome in a series of 32 patients with 2-methylacetoacetyl-coenzyme A thiolase (MAT) deficiency. *Mol Genet Metab* **122**, 67–75 (2017).
101. Drenth, J., Haagsma, C. & van der Meer, J. Hyperimmunoglobulinemia D and periodic fever syndrome. The clinical spectrum in a series of 50 patients. International Hyper-IgD Study Group - PubMed. *Medicine (Baltimore)* **73**, 133–144 (1994).
102. Hoffmann, G. *et al.* Mevalonic aciduria--an inborn error of cholesterol and nonsterol isoprene biosynthesis. *N Engl J Med* **314**, 1610–4 (1986).
103. Coman, D. *et al.* Squalene Synthase Deficiency: Clinical, Biochemical, and Molecular Characterization of a Defect in Cholesterol Biosynthesis. *Am J Hum Genet* **103**, 125–130 (2018).
104. Besnard, T. *et al.* Biallelic pathogenic variants in the lanosterol synthase gene LSS involved in the cholesterol biosynthesis cause alopecia with intellectual disability, a rare recessive neuroectodermal syndrome. *Genet Med* **21**, 2025–2035 (2019).
105. Hoffmann, K. *et al.* Mutations in the gene encoding the lamin B receptor produce an altered nuclear morphology in granulocytes (Pelger-Huët anomaly). *Nat Genet* **31**, 410–414 (2002).
106. Clayton, P. *et al.* Mutations causing Greenberg dysplasia but not Pelger anomaly uncouple enzymatic from structural functions of a nuclear membrane protein. *Nucleus* **1**, 354–366 (2010).

107. Konig, A., Happle, R., Bornholdt, D., Engel, H. & Grzeschik, K. Mutations in the NSDHL gene, encoding a 3beta-hydroxysteroid dehydrogenase, cause CHILD syndrome - PubMed. *Am J Med Genet* **90**, 339–346 (2000).
108. Istvan, E. S. & Deisenhofer, J. Structural Mechanism for Statin Inhibition of HMG-CoA Reductase. *Science (1979)* **292**, 1160–1164 (2001).
109. Bjarnadottir, O. *et al.* Statin use, HMGCR expression, and breast cancer survival – The Malmö Diet and Cancer Study. *Sci Rep* **10**, (2020).
110. Jiang, S. Y. *et al.* Discovery of a potent HMG-CoA reductase degrader that eliminates statin-induced reductase accumulation and lowers cholesterol. *Nat Commun* **9**, (2018).
111. Kirac, D. *et al.* HMGCR and ApoE mutations may cause different responses to lipid lowering statin therapy. *Cell Mol Biol (Noisy-le-grand)* **63**, 43–48 (2017).
112. Wysocka-Kapcinska, M. *et al.* Functional expression of human HMG-CoA reductase in Saccharomyces cerevisiae: a system to analyse normal and mutated versions of the enzyme in the context of statin treatment. *J Appl Microbiol* **106**, 895–902 (2009).
113. Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y. & Takeda, S. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair (Amst)* **5**, 1021–1029 (2006).
114. Storici, F., Snipe, J. R., Chan, G. K., Gordenin, D. A. & Resnick, M. A. Conservative Repair of a Chromosomal Double-Strand Break by Single-Strand DNA through Two Steps of Annealing. *Mol Cell Biol* **26**, 7645 (2006).
115. Storici, F. & Resnick, M. A. The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast. *Methods Enzymol* **409**, 329–345 (2006).
116. Dicarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* **41**, 4336 (2013).
117. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281 (2013).
118. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819 (2013).
119. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science (1979)* **320**, 362–365 (2008).
120. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science (1979)* **285**, 901–906 (1999).
121. Saito, T. L. *et al.* SCMD: Saccharomyces cerevisiae Morphological Database. *Nucleic Acids Res* **32**, D319–D322 (2004).
122. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* **43**, D789–D798 (2015).
123. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
124. The MGC Project Team *et al.* The completion of the Mammalian Gene Collection. *Genome Res* **19**, 2324–2333 (2009).
125. Lamesch, P. *et al.* hORFeome v3.1: A resource of human open reading frames representing over 10,000 human genes. *Genomics* **89**, 307–315 (2007).

126. Burke, D., Dawson, D. & Stearns, T. Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual. *Cold Springs Harbor Laboratory Press* 179–179 https://books.google.ca/books?id=KUzK8kcQ5vEC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false (2000).
127. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
128. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
129. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
130. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
131. Gu, Z. Complex heatmap visualization. *iMeta* **1**, e43 (2022).
132. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
133. Marijanovic, Z. *et al.* Closing the Gap: Identification of Human 3-Ketosteroid Reductase, the Last Unknown Enzyme of Mammalian Cholesterol Biosynthesis. *Molecular Endocrinology* **17**, 1715–1725 (2003).
134. Akhmetov, A. Yeast as a Platform For Synthetic Biology and Investigation of Evolutionary Hypotheses. (University of Texas at Austin, 2019).
135. Souza, C. M. *et al.* A stable yeast strain efficiently producing cholesterol instead of ergosterol is functional for tryptophan uptake, but not weak organic acid resistance. *Metab Eng* **13**, 555–569 (2011).
136. Mollapour, M. *et al.* Yeast Functional Analysis Report Screening the yeast deletant mutant collection for hypersensitivity and hyper-resistance to sorbate, a weak organic acid food preservative. *Yeast* **21**, 927–946 (2004).
137. Chinen, T., Ota, Y., Nagumo, Y., Masumoto, H. & Usui, T. Construction of Multidrug-Sensitive Yeast with High Sporulation Efficiency. *Biosci. Biotechnol. Biochem* **75**, 1588–1593 (2011).
138. Caldas, H. & Herman, G. E. NSDHL, an enzyme involved in cholesterol biosynthesis, traffics through the Golgi and accumulates on ER membranes and on the surface of lipid droplets. *Hum Mol Genet* **12**, 2981–2991 (2003).
139. Villaverde, A., Corchero, J., Sera-Franzoso, J. & Garcia-Fruitos, E. Functional protein aggregates: just the tip of the iceberg-. *Nanomedicine* **10**, 2881–2891 (2015).
140. Narayanaswamy, R. *et al.* Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc Natl Acad Sci U S A* **106**, 10147–10152 (2009).
141. Hu, Z. *et al.* Recent Advances in Ergosterol Biosynthesis and Regulation Mechanisms in Saccharomyces cerevisiae. *Indian J Microbiol* **57**, 270 (2017).

Supplementary Information

Table S1: Primers to confirm presence and absence of human and yeast genes by PCR and sequencing

Gene	Type	Name	Sequence
ScERG1/ HsSQLE	Fp	Fp_ScERG1_HsSQLE	CCTATTGTTGTTTGCTTTTCCTTTTCC
	Rp, Hs	Rp_Conf_HsSQLE	GGTTCCTTTTCTGCGCCTCC
	Rp, Sc	Rp_Conf_ScERG1	CCCAGTCACGTTCTACGATAAG
	Rp, seq.	Rp_Seq_ScERG1_HsSQLE	GCTTAATGTTTGACGGTTCCTATCC
ScERG7/ HsLSS	Fp	Fp_ScERG7_HsLSS	GCATGATTAGCAGAGACATATGGC
	Rp, Hs	Rp_Conf_HsLSS	CCTTAAAGTAATTCTTGGTGTCCAGC
	Rp, Sc	Rp_Conf_ScERG7	CCGATTGTCATGAACATGGGTCC
	Rp, seq.	Rp_Seq_ScERG7_HsLSS	GCGTGTAATCACCATAACCAATAGTTAG
ScERG8/ HsPMVK	Fp	Fp_ScERG8_HsPMVK	GCGATGCGTTTTTCTATACGAGATC
	Rp, Hs	Rp_Conf_HsPMVK	GCCATGCTCCTGAGCATACTG
	Rp, Sc	Rp_Conf_ScERG8	GCACTGAAGGCTCTCAACTCTG
	Rp, seq.	Rp_Seq_ScERG8_HsPMVK	GCGCTGTCACTTCAAACTTCACAT
ScERG8/ HsFDFT1	Fp	Fp_ScERG9_HsFDFT1	GCCCATCTTCAACAACAATACCG
	Rp, Hs	Rp_Conf_HsFDFT1	GCTGCGAAACTGCGACTGGT
	Rp, Sc	Rp_Conf_ScERG9	GCTTCAAAGCTGCCTTCATCTCG
	Rp, seq.	Rp_Seq_ScERG9_HsFDFT1	GGCAAGAAACAGCGTTTACTGG
ScERG10/ HsACAT1	Fp	Fp_ScERG10_HsACAT1	CGTTCGTGGCCAGGTTGATAG
	Rp, Hs	Rp_Conf_HsACAT1	GCTTAGTGGCTGGCAGCAAG
	Rp, Sc	Rp_Conf_ScERG10	CCACTGCTGTCTTGGAGGATAG
	Rp, seq.	Rp_Seq_ScERG10_HsACAT1	GCACACCTCAGCGTTTGAGTAC
ScERG11/ HsCYP51A1	Fp	Fp_ScERG11_HsCYP51A	CGTTTACGTACTCGCATGTATTCG
	Rp, Hs	Rp_Conf_HsCYP51A	GCATGGACAAGAGGTTGCCG

Gene	Type	Name	Sequence
	Rp, Sc	Rp_Conf_ScERG11	CCAACGATTGACTTGGTAGCAGAC
	Rp, seq.	Rp_Seq_ScERG11_HsCYP51A	GGATGAGTGTGGCGATAAGGC
ScERG12/ HsMVK	Fp	Fp_ScERG12_HsMVK	ACAATTTTCGGGCCTCGTTTGGC
	Rp, Hs	Rp_Conf_HsMVK	GGGGTTGAAGCCGGAGGAAT
	Rp, Sc	Rp_Conf_ScERG12	CGGTGGCTTGTTGAGCCTTG
	Rp, seq.	Rp_Seq_ScERG12_HsMVK	GTCTTCGAAAGCTGAATTGATACTACG
ScERG13/ HsHMGCS1	Fp	Fp_ScERG13_HsHMGCS1	GGCGACGATCAAGTCTTCGTC
	Rp, Hs	Rp_Conf_HsHMGCS1	GGGCAACAATTCCCACATCTTTTG
	Rp, Sc	Rp_Conf_ScERG13	GCTCAGATTGGTTGACACATTGAG
	Rp, seq.	Rp_Seq_ScERG13_HsHMGCS1	GCTGGTTCCGCATTTTGTCAAC
ScERG20/ HsFDPS	Fp	Fp_SCERG20_HsFDPS	CGAAGTCAGCTTCTTCTCGTCG
	Rp, Hs	Rp_Conf_HsFDPS	CCCATCTCATCCTCAGTCAGC
	Rp, Sc	Rp_Conf_SCERG20	GGACAAACCTCTATTTAGCTTACCG
	Rp, seq.	Rp_Seq_SCERG20_HsFDPS	CTAGTCACGTGGAACGGTAGATC
ScERG24/ HsLBR	Fp	Fp_ScERG24_HsLBR	CGTACTCCGTCTGAGCATCAAG
	Rp, Hs	Rp_Conf_HsLBR	CCCAGGCCATCGACCTCTTAC
	Rp, Sc	Rp_Conf_ScERG24	CGATTGCCCAGATAGTAGCGC
	Rp, seq.	Rp_Seq_ScERG24_HsLBR	GGAGCGTTGCATAGATAGACCAC
ScERG25/ HsSC4MOL	Fp	Fp_ScERG25_HsSC4MOL	GCGTACGCCTGTGTATGCATAC
	Rp, Hs	Rp_Conf_HsSC4MOL	GCTGAACTAAAGATGCTGACACTTTC
	Rp, Sc	Rp_Conf_ScERG25	GGTGGCCAAAACATCATTGTTCATG
	Rp, seq.	Rp_Seq_ScERG25_HsSC4MOL	GGATATGTATACGTGGTAATGAACGTG
ScERG26/ HsNSDHL	Fp	Fp_ScERG26_HsNSDHL	CGGTCCGATAGTCCCTATAGATCG
	Rp, Hs	Rp_Conf_HsNSDHL	GGTACAGATCCTGTCGGCTGC
	Rp, Sc	Rp_Conf_ScERG26	CCAGAACCACCGATAATTAAAACTGAATC

Gene	Type	Name	Sequence
	Rp, seq.	Rp_Seq_ScERG26_HsNSDHL	CGGTGCGTATCAGATATGCTATTTC
ScERG27/ HsHSD17B7	Fp	Fp_ScERG27_HsHSD17B7	CGCACACCCGTTGTTTTATCTC
	Rp, Hs	Rp_Conf_HsHSD17B7	GCAGTGATCTTATCACCCTGGG
	Rp, Sc	Rp_Conf_ScERG27	CGTTCAAGACACTCACCATGTTG
	Rp, seq.	Rp_Seq_ScERG27_HsHSD17B7	GGCAATGCCTTTTTTGCTTCAC
ScHMG1/ HsHMGCR	Fp	Fp_ScHMG1_HsHMGCR	CTTGAAAGAGCTATATTCGTCTTCGG
	Rp, Hs	Rp_Conf_HsHMGCR	CATGTTCATGGACATCATGCAGATG
	Rp, Sc	Rp_Conf_ScHMG1	CAGTCCCTTGAATAGCGGCG
	Rp, seq.	Rp_Seq_ScHMG1_HsHMGCR	GGTGCTGTTGTGCTTCTTTTTCAAG
ScIDI1/ HsIDI1	Fp	Fp_ScIDI1_HsIDI1	GTCGATGGGGGTTGCCTTTC
	Rp, Hs	Rp_Conf_HsIDI1	GCTGTAGCAGAAGCTTATTTTCGG
	Rp, Sc	Rp_Conf_ScIDI1	GCGTAACTAGATACTGCACCATG
	Rp, seq.	Rp_Seq_ScIDI1_HsIDI1	GGGAAACATTCAAGAGGCCAATAAC
ScMVD1/ HsMVD	Fp	Fp_ScMVD1_HsMVD	CGATGCCGAAATCGCCGAAATG
	Rp, Hs	Rp_Conf_HsMVD	GGCAGAACCAGCTCTTCATCG
	Rp, Sc	Rp_Conf_ScMVD1	CCACAAAGTGTCGCGTTCAAACTC
	Rp, seq.	Rp_Seq_ScMVD1_HsMVD	GCTTGGTCGTTCCGACCAATAC

Table S2: Primers for amplifying repair template with 80bp homology arms

Yeast Gene Homology	Human Gene	Name	Sequence
ERG7	LSS	Fp_ERG7_LSS_100	GCACAAGCCTCTCCAGTAATGTACTGC TGTGCCCAATAACCTTACCAATAATCG TCGCCCACAAAGAAAGTACAAAACAG**A TGACGGAGGGCACGTGTCT**
		Rp_ERG7_LSS_100	TACTACATATACAAATAAATATACTAG TTTTGCACTAGTTTCTAATTGTTGCAG CCTCTAACAACACTTATAAATAAAAC**T CAGGGGTGGCCAGCAAGGG**
ERG9	FDFT1	Fp_ERG9_FDFT1_100	AAGAATAAGAGCACAGAAGAAGAGAAA AGACGAAGAGCAGAAGCGGAAAACGTA TACACGTCACATATCACACACACACA**A TGGAGTTCGTGAAATGCCTCG**
		Rp_ERG9_FDFT1_100	TTTAAGGATAGGCCTCTACCTATTATG TAAGTACTTAGTTATTGTTCGGAGTTG TTTGTTTATGTTATTTGGCGCAGACT**T CAGTGTTCTCCAGTCTGAACATA**
ERG10	ACAT1	Fp_ERG10_ACAT1_100	CGTTTTCTAGTTCTCTGAAAAAAGGTA GCCTAAAACAAGCGCCATATCATATAT ATTTATACAGATTAGACGTACTCAAA**A TGGCTGTGCTGGCGGCACT**
		Rp_ERG10_ACAT1_100	AATTATGCCTGTTATACATAAAGCCAT TTATATATTTATGTATTTTATGAAAAA GATCATGAGAAAATCGCAGAACGTAAC **TACAGCTTCTGAATTAGCATGGCA**
ERG11	CYP51A1	Fp_ERG11_CYP51A1_100	AATAGTTTCCAGAAAAAATTTTTTTTC CTTCACAATTGCAGCAGGCTTGAATAG AAACAGAACAAACGAGTAATACAAGG**A TGGCGGCGGCGGCTGGGAT**
		Rp_ERG11_CYP51A1_100	GATGAGTGTGGCGATAAGGCCCACGTG AATGTACAACTTCTCTCTTTTTCTGTT TTTTTTTTTTTCTCAGTTACAAACC**T CATTTTGATCTTCGTTTGTAACGGATA AC**
ERG12	MVK	Fp_ERG12_MVK_100	ATGATATTACTTCCTGGTTAGCTGGTG CGCTTCTTAGTCCTAACTTGCAAATTT ATATCTACGTATAGAAAACTGTCAAT**A TGTTGTCAGAAGTCCTACTGGTG**

		Rp_ERG12_MVK_100	CTTTCTATCTTTATTACATAAAACTTC ATACACAGTTAAGATTAAAAACAACTA ATAAATAATGCCTATCGCAAATTAGC**T** **CAGAGGCCATCCAGGGCTT**
ERG13	HMGCS1	Fp_ERG13_HMGCS1_100	GGCTCCATTCGGCCTCATCGGGCGGAA CTGTTTGGGTGCCTCGCCAATTTTCGT AGCCAAATTACTATAAAAGGTGCAGC**A** **TGCCTGGATCACTTCCTTTGAATG**
		Rp_ERG13_HMGCS1_100	GAGAGCTAAAAAAAGCAGGCTGCCAAA AAAATAAAGCATTTATGAAGGGGGTTC AGCAAGATGCAATCGATGGGGGAAGAT **TAATGTTCCCCATTACTAATGACAGC**
ERG20	FDPS	Fp_ERG20_FDPS_100	TTTTCATTTTAACTTTAAAAACTCAAC CAACAGGTATTGGACTGACATAGGCAC AATAAACTCAAAAATATTACGTAGAA**A** **TGCCCCTGTCCCGCTGGTT**
		Rp_ERG20_FDPS_100	TATTATATTGGAAGTATATAGTTCTTA TACATGGTCCTTATCTAGTTTGAAATC TAATGTTTTATCGATTAGCGTTAGTT**T** **CACTTTCTCCGCTTGTAGATTTTG**
ERG25	SC4MOL	Fp_ERG25_SC4MOL_100	TTAGTTGTAACTTTTTCTCTTTAGATA GTAGCATAGAGGACTAAGGAAAAGTAG TACAGCCATAAAAAAAAGAGGAAAAG**A** **TGGCAACAAATGAAAGTGTCAGCATC**
		Rp_ERG25_SC4MOL_100	AAAGCGGTTCGAAATTTATTTCATTTT ATACGTTTTTTTGTTGTTTTTCTTTTT TTTTTTTTTGAAGTATGTTTCTTCTCT TAT**TCAGTCTTTTTCTCAAACTTCTTC** **C**
EGR26	NSDHL	Fp_ERG26_NSDHL_100	TAGAATAGACAACAATATAGATTGTCA ATTTTGTCCAATTTTATTTGCAACTCT ACCTGGAAGGGCAAACAAAAATAAAT**A** **TGGAACCAGCAGTTAGCGAG**
		Rp_ERG26_NSDHL_100	TATATACGGTGCGTATCAGATATGCTA TTTCAACATACTATCTTTCGATAATCG GATCAAAAAGCTCCTAACGATTGCCA**T** **CACTTGACCCTCCGCAGGT**
ERG27	HSD17B7	Fp_ERG27_HSD17B7_100	GTATTTAAAAGATAGGACGAGAAACAA GCACATGATCTGTGTCGAAAAAAAGTA GCAAAGAGAAAAAGTAGGAGGATAGG**A** **TGCGAAAGGTGGTTTTGATCACC**

		Rp_ERG27_HSD17B7_100	ATCAGTTGATATGCGTTTTCTATACAG ACATACATGTATATACACACATATATA TATATACATATGTACGCAGAGATATA**T** **TATAGGCATGAGCCACTGAGC**
MVD1	MVD	Fp_MVD1_MVD_100	GATTGATATATAAAGAGCATCCATCGT GCTGGTAATGGTGATGCAAGATTGAAA ATTCCCAGACAGTGAGCACCAGCACA**A** **TGGCCTCGGAGAAGCCGCT**
		Rp_MVD1_MVD_100	ATACCGTAGGGAAACGTAAATTAGCTA TTGTAAAAAAAGGAAAAGAAAAGAAAA GAAAAATGTTACATATCGAATTGATC**T** **CAGGCAGCTGGCTTCGGCA**

Table S3: Human gene coding sequence templates

Gene	Source	Entrez	Template length	Accession ID
ACAT1	IDT	38	1284	NM_000019.4
HMGCS1	Horizon MGC	3157	1563	NM_001098272.3
MVK	Human ORFeome	4598	1191	NM_000431.4
MVD	Horizon MGC	4597	1203	NM_002461.3
FDPS	Human ORFeome	2224	1260	NM_002004.4
FDFT1	Human ORFeome	2222	1254	NM_004462.5
LSS	Human ORFeome	4047	2199	NM_002340.6
CYP51A1	Horizon MGC	1595	1530	NM_000786.4
SC4MOL	Human ORFeome	6307	882	NM_006745.5
NSDHL	Human ORFeome	50814	1122	NM_015922.3
HSD17B7	Human ORFeome	51478	1026	NM_016371.4

Table S4: Unique SNPs from WGS of 5-gene humanised alpha proteasome strain

Chromosome	Start	Δ	Type	Variant Freq.	Variant P-Value	Amino Acid Δ	CDS	Protein Effect
4	987244	G -> A	SNP	1	$1x10^{-154}$			
7	38443	G -> A	SNP	1	0	S -> L	RAI1	Subst.
7	762093	T -> C	SNP	1	$1.6x10^{-281}$		HsPSMA4	None
12	554	-T	Deletion	1	5×10^{-8}			
12	1901	C -> T	SNP	0.993	$4.6x10^{-42}$	D -> N	Y' element ATP-dep. helicase CDS	Subst.
12	216700	G -> A	SNP	1	$1.6E^{-14}$	S -> L	gag-pol fusion	Subst.
12	518361	G -> C	SNP	1	$2.5E^{-256}$	R -> S	SWI6	Subst.
13	820490	G -> A	SNP	1	$6.3E^{-314}$	V -> I	FCP1	Subst.
13	921381	(T)14 -> (T)13	Deletion	0.961	$2E^{-105}$			
15	114980	T -> C	SNP	1	0	T -> A	WSC3	Subst.
15	255822	T -> C	SNP	1	$1.6E^{-296}$		HsPSMA7	None
15	255843	C -> T	SNP	1	$4.0E^{-321}$		HsPSMA7	None
15	548683	T -> C	SNP	1	$1E^{-288}$			
15	881948	C -> T	SNP	1	$4E^{-303}$	H -> Y	RAX1	Subst.
15	1017930	C -> T	SNP	0.988	$3.4E^{-301}$		HsPSMA3	None
15	1018167	T -> C	SNP	1	$1E^{-294}$	I -> M	HsPSMA3	Subst.
15	1018173	T -> C	SNP	1	$1E^{-294}$		HsPSMA3	None

Table S5: ScEDIT collection sgRNA sequences cloned into CRISPR plasmids

Pathway	Target locus	sgRNA sequence	Creator
Proteasome core	ScSCL1	CATCACTATCTTTTCCCCCG ATAAACTCACTAGCGGTCAG	Mudabir Abdullah
	ScPRE9	ACAACAATTTTCTCCCCTGA TTGGGATTATGGCATCTGAT	
	ScPRE6	AGATGGACACATTTTCCAAG AGTGGAGTACGCCCTGGAGG	
	ScPUP2	ACTAGAAGTGAATATGATCG GCACATTTTCCCCAGAAGGG	
	ScPRE5	GTTCAGGAACAATTACGACG GTTCCAAGTGGAATACGCCT	
	ScPRE10	ATCGGTATAAAGTGTAACGA CAAGTCGTAGACCGTCACAT	
Proteasome base and lid	ScRPT2	CCAAATTTAGATTGAACGGG	Brittany Greco
	ScRPT3	GATGCGTATGATTTGACGGC	
	ScRPT5	AGACGGAAAGTGAAACGTGA	
	ScRPT6	ACAACGTTAGCCACAAGGTA	
	ScRPN2	GTATTAGAAACCCACGAAAG	
Safe harbour site	511b	CAGTGTATGCCAGTCAGCCA	Leanne Bourgeois
	Userx-1	CAGTGTATGCCAGTCAGCCA	
	Fgf20	GTTAGAGCTGTTACAAGTTA	
	Fgf24	CCTATTGGACAAGATTTACG	
Adenine synthesis	ScAde2		Brittany Greco
Sterol biosynthesis	ScERG10	CTAGCACCAGAACAACCCAA	

	ScERG13	GGTGCCGCAAGACCAACCGG	
	ScERG9	GGCAATGTAAAGATTCGTAA	
	ScERG24	TGATGGATATCACTACAGAT	
	ScERG27	ATTCTGCCTCAATTGACCAG	
	ScERG1	GTAGCCGTAAACTATATCGC	
	ScERG9	GATCTTACCGATCTGTTCTG	
	ScERG7	GGTGAATTGATGCAACCAGG	Aashiq Kachroo
	ScERG11	TTACGGAACAAAGAGTGCAA	
	ScERG12	TTAAAGTCTACTTTACCCAT	
	ScERG25	GCTCACCGTCTATTCCACTA	
	ScHMG1	CTACTGATGGAACGAAATGG	
	ScERG26	TATGACATAGTGAATGTTAA	
	ScERG8	CATGCTGTASCCCATCCTTA	
	ScIDI1	AATTCGTCATCAATACATAG	
	ScMVD1	ACGCCAATTAAGAAAGGAAA	Michelle Vandeloo

Table S6: Golden Gate primers for assembly of repair template in LEU2 plasmid YTK97.

Gene	Part	Name	Sequence
ScERG25/ HsSC4MOL	HsSC4MOL + 5'/3' 300bp homology arms (from HsS1.7)	Fp_GG_SC4MOL_Erg25_BsmBI	CACCACACGTCTCGCTGATC TGCTGTGGTCGTATTATT
		Rp_GG_SC4MOL_Erg25_BsmBI	CACCACACGTCTCGTGCTCG GTTGAAAGATTTTTGTAGT
ScERG27/ HsHSD17B7	300bp 5'UTR of ScERG27	Fp_GG_5UTR_ERG27_BsmBI	CACCACACGTCTCGCTGAAA GCTGCTCAACTGGGTCC
		Rp_GG_5UTR_ERG27_BsmBI	CACCACACGTCTCGCCTATC CTCCTACTTTTTCTCTTTGC
	HsHSD17B7 coding sequence	Fp_GG_HSD17B7_BsmBI	CACCACACGTCTCGTAGGAT GCGAAAGGTGGTTTTGATCA
		Rp_GG_HSD17B7_BsmBI	CACCACACGTCTCGTTATAG GCATGAGCCACTGAGC
	300bp 3'UTR of ScERG27	Fp_GG_3UTR_ERG27_BsmBI	CACCACACGTCTCGATAATA TATCTCTGCGTACATATGT
		Rp_GG_3UTR_ERG27_BsmBI	CACCACACGTCTCGTGCTTT TCTGGAACGACTCAGGAT

Table S7: Amino acid sequence identity alignment between yeast and human orthologs of the sterol biosynthesis pathway.

Yeast gene	Human ortholog	Amino acid sequence identity
ERG10	ACAT1	48.241%
ERG13	HMGCS1	47.046%
HMG1	HMGCR	32.468%
HMG2	HMGCR	31.754%
ERG12	MVK	32.151%
ERG8	PMVK	8.609%
MVD1	MVD	46.250%
IDI1	IDI1	48.319%
ERG20	FDPS	44.034%
ERG9	FDFT1	40.845%
ERG1	SQLE	35.070%
ERG7	LSS	40.515%
ERG11	CYP51A1	36.364%
ERG24	LBR	38.652%
ERG25	SC4MOL	36.789%
ERG26	NSDHL	36.827%
ERG27	HSD17B7	22.063%

Figure S1: Whole genome sequencing reveals precise replacement of multiple genes in a single strain

Using a reference genome with the human (Hs) genes replaced at the native loci of the yeast orthologs, WGS shows good coverage and accurate insertion of the five human proteasome genes in yeast. Sequence alignment, coverage analysis, and SNP discovery was done in Geneious. Numbered axis above all sequences represents position on the chromosome. Green and yellow arrows represent coding sequences/genes. Orange arrows indicated inserted human gene sequence, differing from the native yeast gene sequence. Green bars represent mapped reads of expected quality and length (400-500bp) (blue-tinted are longer than expected, yellow-tinted are shorter than expected). Purple bars represent unpaired reads.

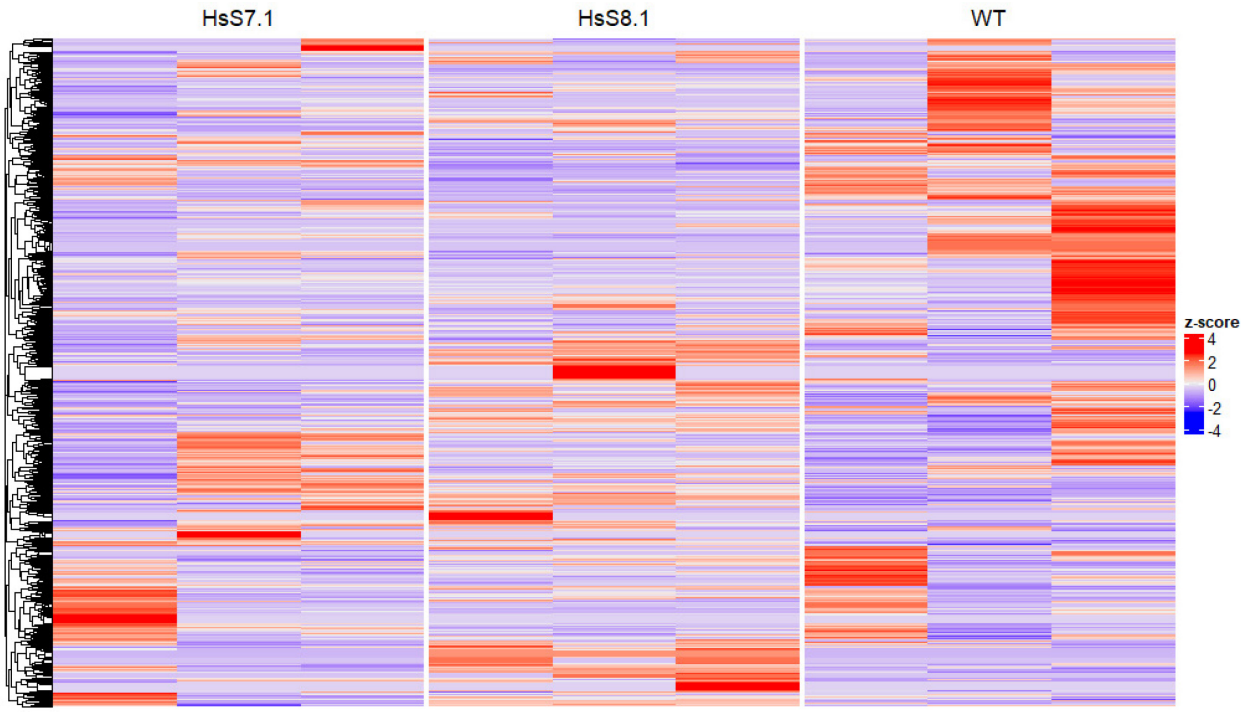


Figure S2: Clustered heatmap of differentially expressed proteins in humanized yeast.

Z-scores were computed based on unique spectrum counts from whole cell protein lysate after LC/MS across groups for all yeast proteins expressed in two humanised strains (HsS7.1 and HsS8.1) compared to the wild-type (WT). The colour scale describes relative shifts in protein expression between samples (red = increased, blue = decreased). Each replicate is its own column, and each sample was run in triplicate. Clustering was done by z-score similarity, independent of protein function or annotation.