



Cite this: *Lab Chip*, 2023, 23, 3704

Large scale microfluidic CRISPR screening for increased amylase secretion in yeast†

S. Andreas Johansson, ^a Thierry Dulermo,^b Cosimo Jann,^{‡a} Justin D. Smith, ^{§cd} Anna Pryszlak,^{¶a} Georges Pignede, ^b Daniel Schraivogel, ^a Didier Colavizza,^b Thomas Desfougères,^b Christophe Rave,^b Alexander Farwick, ^b Christoph A. Merten, ^a Kevin R. Roy,^{cd} Wu Wei ^{**††cd} and Lars M. Steinmetz ^{*acd}

Key to our ability to increase recombinant protein production through secretion is a better understanding of the pathways that interact to translate, process and export mature proteins to the surrounding environment, including the supporting cellular machinery that supplies necessary energy and building blocks. By combining droplet microfluidic screening with large-scale CRISPR libraries that perturb the expression of the majority of coding and non-coding genes in *S. cerevisiae*, we identified 345 genes for which an increase or decrease in gene expression resulted in increased secretion of α -amylase. Our results show that modulating the expression of genes involved in the trafficking of vesicles, endosome to Golgi transport, the phagophore assembly site, the cell cycle and energy supply improve α -amylase secretion. Besides protein-coding genes, we also find multiple long non-coding RNAs enriched in the vicinity of genes associated with endosomal, Golgi and vacuolar processes. We validated our results by overexpressing or deleting selected genes, which resulted in significant improvements in α -amylase secretion. The advantages, in terms of precision and speed, inherent to CRISPR based perturbations, enables iterative testing of new strains for increased protein secretion.

Received 7th February 2023,
Accepted 7th July 2023

DOI: 10.1039/d3lc00111c

rsc.li/loc

Introduction

The baker's yeast, *Saccharomyces cerevisiae*, has been extensively used to study the secretory pathway¹ and autophagy,² with

applications in both biotechnology and clinical research. Yeast is used to produce high-value products, including pharmaceuticals, antibiotics, antioxidants, vitamins, and recombinant proteins such as insulins, albumins, and antibodies.³ The commercial value of recombinant proteins is predicted to double by 2025,⁴ emphasising the current and long-term potential of yeast and other microorganisms as cell factories. Further improvements in the efficiency and yield of recombinant protein production and secretion, will help to reduce the cost and environmental footprint of this industry. An improved understanding of the processes involved in protein secretion also has implications for the treatment of several human diseases in which misfolding of proteins results in aggregate formation and cellular dysfunction.^{5,6}

Proteins destined for secretion in eukaryotic organisms are processed through the ER in which nascent proteins are folded and glycosylated, to the Golgi apparatus for further modifications, and packaged into secretory vesicles for delivery to the plasma membrane. The complexity of these processes involving a large number of components and pathways, coupled with system wide metabolic demands for energy and building blocks, suggest that large-scale screens are most suited for the identification of bottlenecks and targets for improvements. Systematic studies of the effects on protein secretion from gene perturbations are challenging.

^a European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. E-mail: larsms@embl.de

^b Lesaffre Institute of Science & Technology, Lesaffre, 59700 Marcq-en-Baroeul, France

^c Department of Genetics, Stanford University School of Medicine, Stanford, California, USA

^d Stanford Genome Technology Center, Stanford University, Palo Alto, California, USA

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3lc00111c>

‡ Present address: Biocentre, Departments of Biology and Chemistry, Johannes Gutenberg University Mainz, Mainz, Germany & Institute of Molecular Biology gGmbH, Mainz, Germany.

§ Present address: Recombia Biosciences, 1000 Marina Blvd STE 450, Brisbane, CA 94005, USA.

¶ Present address: VERAXA Biotech GmbH, 69124 Heidelberg, Germany.

|| Present address: Institute of Bioengineering, School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

** Present address: CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China.

†† Present address: Center for Biomedical Informatics, Shanghai Engineering Research Center for Big Data in Pediatric Precision Medicine, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China.



The secreted protein of interest must be captured for quantification using a method that is amenable to pooled screening. Several recent studies have used microfluidic platforms to identify oversecretion strains in cell populations perturbed by mutagenesis or RNA interference.^{7–9} However, whole genome sequencing is needed to identify the affected locus or loci in mutagenised cells, while RNAi is limited to repressional interference of gene expression, with implications for the speed and scope of discovery.

Here, we perform a systematic interrogation of the effects of gene activation and repression on α -amylase secretion across more than 7000 genes and ncRNAs in yeast. Using large-scale CRISPRi¹⁰ (repression) and CRISPRa¹¹ (activation) libraries, we systematically probe the effects from perturbations of gene expression on α -amylase secretion. The application of CRISPR/Cas9¹² in combination with high throughput screening and next-generation sequencing (NGS) allowed us to maintain a near genome-wide scope with single gene precision. We identified an increased expression of genes involved in vesicle trafficking, the late Golgi, including retrograde transport from the endosome, and repression of expression of genes involved in the cell cycle and ribosomal processes as beneficial for increased α -amylase secretion. Our results can be applied for yeast strain engineering, to facilitate protein secretion in biotechnological applications, and to improve our understanding of protein processing and secretion in general.

Methods

Selection of guide-RNA and oligo design

A single library of guide RNAs for both CRISPRi and CRISPRa, was designed to target promoter regions of 7032 ORFs, SUTs (stable unannotated transcripts), and CUTs (cryptic unstable transcripts), for simplicity, referred to as genes from here on, based on previously published transcription boundaries in *S. cerevisiae*.¹³ For each gene, we identified all potential guide RNAs (gRNA) with minimal predicted off-target effects from 400-bp upstream to 100-bp downstream of the transcription start site (TSS). If there were less than 6 gRNAs in one gene promoter region, we selected all available. Otherwise, we divided the promoter regions into six 50-bp bins, from 300-bp upstream to the TSS. From each bin, the SpCas9 guide with the highest scoring was selected according to the Azimuth prediction software.¹⁴ This resulted in a library of 40 890 guides (Table S1†) with an average of approximately six guides per feature. The potential for off-target effects was minimised by blasting the individual guide RNAs against each other, as well as all potential gRNA binding sites (4.7 M in total) throughout the genome and removing any guide with less than three mismatches. Oligos were ordered from Agilent using a design that optimises the number of guides per oligo: each 190 bp oligo contains four individual 20 bp guide-RNA sequences interspersed with spacer sequences containing double type II-S recognition sites, enabling restriction digest and release using BspQI with subsequent removal of the

recognition site. Example: TCAGTCGATCGgtcttcaaggAAGATA TACGTTATTGATATgttagaagagcgtcttctaggGGAAGGAATATTGA GCAACAgttagaagagcgtcttctaggGCGGGTAACGACAACGAAG TgttagaagagcgtcttctaggTCTCGATTACCAAACCT TgttgaagagcGCTAGCTCCAT.

Plasmid vectors and background strain

Two plasmids with backbones previously derived¹⁵ with nuclease-null dCas9 fused to the tripartite VP64-p65-Rta (VPR) domain¹¹ for gene activation or the Mxi1 domain¹⁶ for gene repression were developed for this project. The active domains together with the dCas9 are expressed using a TEF1 constitutive promoter for continuous production of the resulting fusion protein, while the expression of the gRNA is inducible using a TET-On system upon activation by addition of anhydrotetracycline. Selection for the plasmid was maintained using a *kanMX* cassette. Codon optimised α -amylase (AnAmy6) was successfully integrated into the genome of the *S. cerevisiae*, commercially available Ethanol Red background strain at the HO locus with or without Cas9 assisted cutting of the diploid wild-type allele and made homozygous by subsequent loss of heterozygosity. The cassette was integrated with different promoters using a hygromycin selectable marker in two or more copies as estimated using qPCR.

Plasmid assembly and cloning

Oligos were PCR-amplified using KAPA HiFi polymerase for 20 cycles and the guides released using BspQI restriction digest overnight in CutSmart buffer at 50 °C, followed by Antarctic phosphatase treatment at 37 °C for 30 min and heat inactivation at 80 °C for 2 min. Subsequent reannealing was performed at room temperature. Plasmid vectors were similarly digested overnight at 50 °C using BspQI, followed by a second digest of the released spacer using AscI for the activation plasmid and NotI-HF for the repression plasmid, and clean-up using a Monarch PCR & DNA cleanup kit (New England Biolabs, MA, USA). DNA inserts encoding guide RNAs (10 ng) were mixed with plasmid vectors (0.5 to 1 μ g) and concentrated T4 DNA ligase (2 million units per μ l) in T4 DNA ligase buffer and incubated at 25 °C for 1 hour before heat inactivation at 65 °C for 10 min. The ligation product was dialyzed for 25 min on Type-VS Millipore membrane against distilled water. Plasmids were transformed using electroporation at 1.8 or 1.9 kV, 200 Ohm, 25 μ F into NEB 10-beta electrocompetent *E. coli* cells following the manufacturer's protocol using 2 μ l of ligation product per 25 μ l of cells. The cells were spread on LB plates with kanamycin added at 30 μ g ml⁻¹ and incubated overnight after 1 hour of outgrowth at 37 °C. Following outgrowth, the cells were scraped off the plates into LB on ice and centrifuged to collect the cell pellets. The cloned plasmid libraries were extracted using the QIAGEN Plasmid Maxi kit and stored at -80 °C.



Yeast transformation and harvesting

The two resulting plasmid libraries (activation or repression) were transformed separately into the background strain using a modified protocol based on Benatui *et al.*¹⁷ with the addition of 100 μ l denatured herring sperm (2 mg ml⁻¹) to 400 μ l of cells in electroporation reaction buffer to further increase the efficiency. Each reaction contains 400 μ l of cells with 5 μ g of plasmid libraries. Electroporation was performed at 2.5 kV, 400 ohm, and 25 μ F, followed by immediate

addition of 1 mL of YPAD and outgrowth for 2 hours at 30 °C at 450 RPM on a shaking heater block. After centrifugation for 2 min at 7000 RPM, the cells were resuspended in 1 M sorbitol and plated on YPAD with sorbitol (1 M) and geneticin (400 μ g μ l⁻¹). The yeast colonies containing the final transformed libraries were collected the next day using a glass cell spreader and the addition of 2 ml of YPAD. The suspended cells were transferred to an Erlenmeyer flask on ice, followed by centrifugation in 50 ml Falcon tubes at 4 °C to pellet the cells. Collected cells were resuspended in 500 μ l

A CRISPR libraries

Microfluidic screen

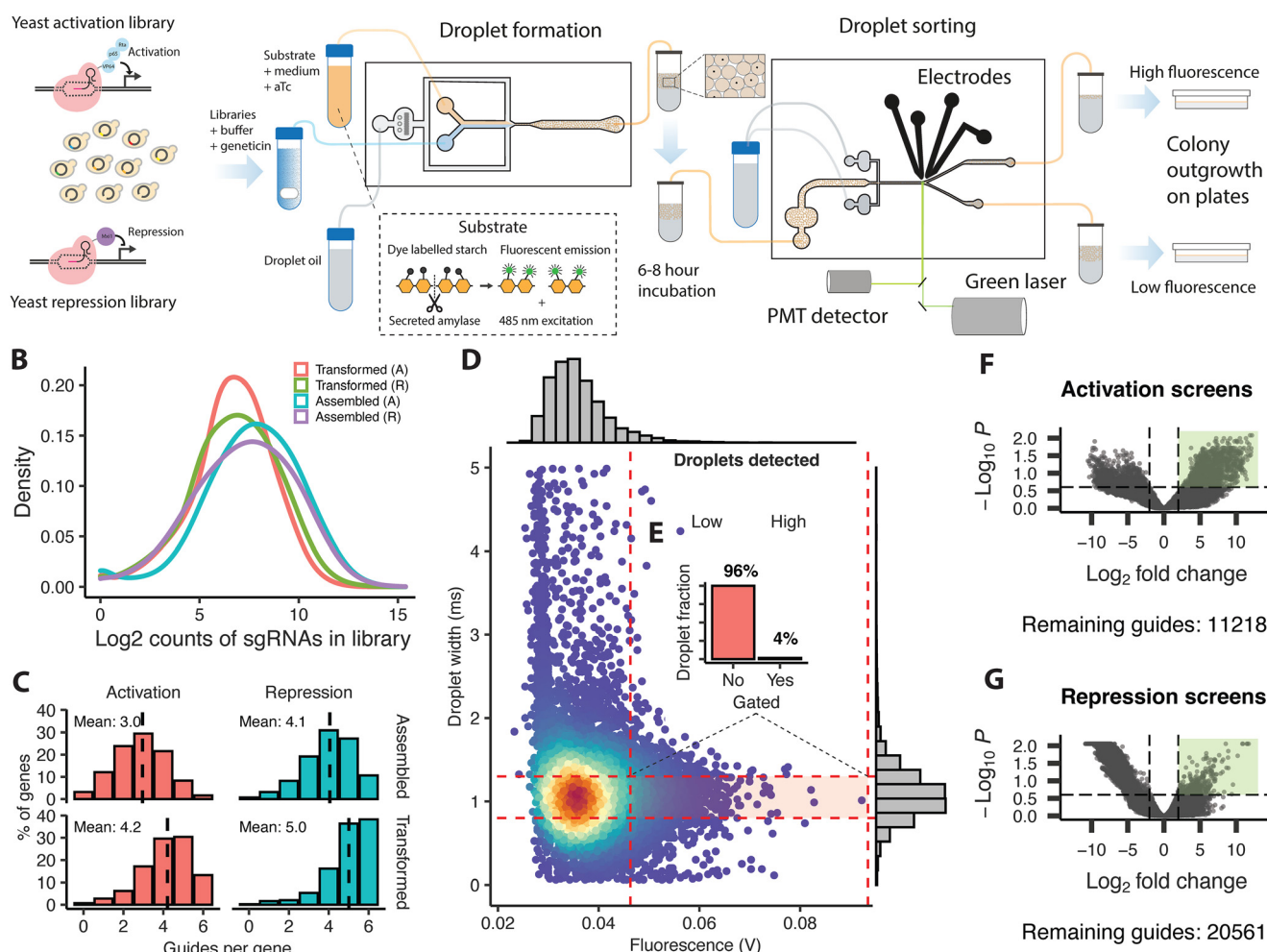


Fig. 1 Schematic illustration of the CRISPR based perturbation and microfluidic screening of yeast cells for increased α -amylase secretion, and overview of the assembled libraries, screening, and sequencing of sorted cells. (A) Plasmid-based gene activation and repression libraries were transformed into α -amylase secreting yeast. These libraries were mixed with a fluorescent substrate, growth medium and anhydrotetracycline to induce gene activation or repression and loaded on a droplet forming chip to generate droplets. The droplets were incubated for 6–8 hours at ambient temperature to allow for protein secretion and then transferred to a droplet sorting chip. The droplets were sorted into a high and low fluorescence fraction using electrodes based on the droplet width and the amount of emission from the fluorescent substrate from a single droplet. Sorted cells were released from the droplets and plated for outgrowth before processing for sequencing. (B) Density distribution of guide counts per library during assembly and transformation for activation (A) and repression (R) libraries, (C) histogram showing the average number of surviving guides per gene and the percentage fraction of the total guide pool during assembly and transformation for activation and repression libraries, (D) microfluidic sorting example showing the ratios (E) for the gated fraction (high fluorescence) versus the non-gated population (low fluorescence) fraction based on the droplet width and fluorescent emission during a single round of sorting. Volcano plots of enriched guides (green rectangle) as identified by NGS for the activation (F) and repression screens (G) and the number of remaining guides meeting the filtering criteria, from the statistical analysis comparing counts of guides from the low and high fluorescence fractions from replicate rounds of screening.



of YPAD and mixed with 30% glycerol before storage at -80°C . Sequencing of the original assembled and transformed libraries identified a gRNA representation of 72 and 86%, respectively, for the activation and the repression libraries following assembly, and 49 and 69% after retransformation into yeast. The two final yeast libraries with activation or repression plasmid libraries had an average coverage of 3.1 and 4.0 guides per gene (Fig. 1C).

Microfluidics

The design and manufacturing of PDMS chips and the equipment used for the microfluidic screening is described in Chaipan *et al.* (2017).¹⁸

Droplet formation

Yeast cultures with plasmid libraries were inoculated using 50 μl of stock culture in 50 ml of YPAD with geneticin for overnight growth. Cells were collected by centrifugation and washed twice in phosphate-buffered saline (PBS), sonicated for four rounds of 10 seconds on ice, and adjusted to 125 million cells in 5 mL PBS with geneticin ($400\text{ }\mu\text{g mL}^{-1}$) to achieve a final cell to droplet ratio of 0.4. Substrate for α -amylase detection from the EnzChek Ultra Amylase assay kit (Invitrogen) was prepared according to manufacturer's instructions and adjusted to 1 mL in PBS before the addition of 4 mL of YPAD and 500 $\text{ng }\mu\text{l}^{-1}$ anhydrotetracycline (aTc). The cells with PBS and the substrate in YPAD were loaded on separate 5 ml syringes, with a small bar magnet added to the cell syringe and a 20 μm PES filter added to the substrate syringe. The latter was further covered in aluminium foil to protect the light-sensitive components.

Droplets were generated on a flow focusing droplet generator chip, using flow rates of 400–500 μl per hour for the aqueous phase, and 1100 μl per hour for the oil phase (QX200 Droplet Generation Oil for EvaGreen, Bio-Rad). Once a stable flow had been established, droplets were collected into 1 ml syringes which were then placed on ice and protected from light. The syringes were sealed with parafilm and aluminium foil and stored overnight at 7°C , then incubated for 6–8 hours at room temperature to allow protein secretion and amylase substrate conversion within the droplets.

Droplet sorting

The droplets with cells were introduced into the sorting chip at a flow rate of 20 μl per hour and spaced using oil infused at 500 μl per hour. The fluorescence emission of single droplets resulting from excitation using a green laser (488 nm excitation) was measured using a photomultiplier tube. The gain of the signal was increased until individual droplets were resolved. Gating was applied to droplets with the highest fluorescent emission, using a sorting window centred on droplets with an average droplet size as shown in Fig. 1E and adjusted to select the top 2–5% of the total population. Single droplets were sorted by application of 1.5 to 1.7 kV

across the electrodes using 3–4 ms pulses of square waves at a frequency of 40 kHz to drive the gated population into a sorting channel.

Harvesting of cells and plasmid extraction

Collected droplets were mixed with 5–10 μl of the droplet destabilising chemical 1H,1H,2H,2H-perfluoro-1-octanol (Sigma) after removal of excess oil and vortexed to break the droplets. The released cells were mixed with a small amount of fresh medium without anhydrotetracycline and streaked on YPAD geneticin plates for overnight growth. Cells were collected from plates in 2 ml of YPAD, pelleted by centrifugation, and processed using a FastPrep vortexing machine with added acid-washed glass beads, followed by plasmid extraction using the QIAprep Spin Miniprep kit (Qiagen).

Guide-RNA amplification and NGS sequencing

The 20-bp variable guide-RNA region on the extracted plasmids was PCR amplified using Q5 high-fidelity DNA polymerase (NEB) and primers specific to the common flanking regions directly outside the guide-RNA insertion site, with added overhanging sequences resulting in the introduction of 4 to 6-bp barcodes followed by Illumina sequencing adapters on the amplified product. The PCR product from multiple reactions per barcode and sample was mixed and cleaned using Ampure XP beads before sequencing on the Illumina NextSeq500 platform using paired-end 75 nt reads.

Bioinformatics

Reads were demultiplexed using Je,¹⁹ aligned using BWA²⁰ against padded guide sequences, and finally summarised into a count table using an R script. Statistical analysis was carried out using edgeR.^{21,22} Guide counts were normalised using the RLE method,²² with counts from each low and high fluorescence fraction from multiple screens (2 in total for the repression libraries and 4 in total for the activation libraries, each with 2 replicates) treated as replicates ($n = 4\text{--}8$). Statistical tests were performed using the glmFit and glmLRT functions which fit generalised linear models and conduct likelihood ratio tests. The Benjamini–Hochberg method²³ was applied to the p -values to control the false discovery rate (FDR). Guides for genes meeting a criteria of [$\log FC > 3$, $\text{FDR} < 0.05$] were identified as significantly enriched. The fold coverage for screening was calculated by multiplying the number of screened droplets with the loading efficiency ($\lambda = 0.4$) and dividing the sum with the original library size.

Enrichment analysis

The guide with the highest fold-change for an individual gene [$\log FC > 3$, $\text{FDR} < 0.3$] was selected as representative and tested for enrichment using R packages gprofiler²⁴ or TopGO²⁵ via ViSEAGO.²⁶ Results from gprofiler2 were FDR corrected for multiple testing errors while TopGO results



from Fisher's exact test were instead pruned using the elimination algorithm. Results from both tests with a p -value below 0.05 were considered significantly enriched. Gene network analysis was performed using ClueGO^{27,28} and including genes with a $\log FC > 8$ from the early activation screen [$FDR < 0.1$] and the main activation screens [$FDR < 0.05$], and genes from the repression screens [$\log FC > 3$, $FDR < 0.15$]. Evidence codes from all available sources were used for the statistical testing using a right-sided hypergeometric test with FDR correction (p -value < 0.05).

Validation of α -amylase secretion

PCR-amplified native candidate genes were cloned into plasmid p427-TEF between SpeI and SalI and transformed into the background strain for plasmid-based overexpression. Deletion strains were constructed by golden gate assembly of annealed oligos with gRNA sequences targeting the start and end position of the target gene, into sgRNA expression vector pWS082. The assembled plasmid and Cas9 expression vector pWS173 were linearized using EcoRV or BsmBI and co-transformed with annealed repair fragments, consisting of the joined 60 bp flanking regions of each target gene, which upon successful homology-directed repair, resulted in the deletion of the target gene. The insertion or deletion of the target gene was confirmed by Sanger sequencing. The amount of secreted amylase was measured in cultures grown in YPAD medium after 4, 24 and 48 hours using the Ceralpha method using α -amylase assay kits from Megazyme, Ireland. The relative amount of secreted α -amylase per cell was calculated by dividing the measured amount of secreted α -amylase by the optical density (OD600) as measured during the point of sampling. Baseline expression was measured using the transformed empty vector, or the background strain without gene deletion, for overexpression and deletion strains respectively.

Statistical analysis

Results reported as significant for the *in vitro* validation are based on the results from the analysis of variance (ANOVA) and the pairwise means from Tukey HSD using the AOV and TukeyHSD functions in R, respectively, together with an interaction model for the two experimental conditions strain and time-point. The normality assumption was tested using the Shapiro-Wilks test, while homogeneity of variances was tested using Levene's or Bartlett's test.

Results

A pooled CRISPR/Cas9 screening platform for protein secretion in yeast

We utilised CRISPR with nuclease-null dCas9²⁹ to perturb a single gene per cell in a pooled format across the genome, coupled with droplet microfluidic sorting (Fig. 1A) using a previously established, fluorescent α -amylase assay^{7,8} and chip designs.¹⁸ The commercially available strain Ethanol

Red, commonly used to produce bioethanol, was used as the background strain.³⁰ We engineered this strain to express α -amylase by insertion of an expression cassette containing the codon-optimised α -amylase gene from *Aspergillus niger* into the HO locus and transformed it with plasmid activation or repression libraries (Fig. 1B and C). The microfluidic system was first used to create droplets containing cells ($\lambda = 0.4$) from the transformed protein-secreting strain, together with the fluorescent substrate, growth medium, and anhydrotetracycline (which induces the expression of the guide RNA). These droplets were incubated off chip to accumulate secreted protein, and then reintroduced on a second microfluidic chip for sorting. Degradation of the substrate by secreted amylase relieves quenching of a BODIPY dye conjugate. Single droplets were assayed for secreted α -amylase content, by measuring the fluorescent emission resulting from laser excitation of the unquenched fluorescent dye. Droplets of average size with the 2–5% highest fluorescence signal (Fig. 1D and E) were sorted into a high fluorescence fraction by application of an electrical field, with the remaining droplets passing passively into a low fluorescence fraction. The cells were released from the droplets in fresh medium without anhydrotetracycline and plated out for overnight recovery and outgrowth, followed by PCR amplification of the plasmid guide region. Next generation sequencing of the amplicons allowed us to quantify gRNA abundance, and identify enriched genes, by comparing the guide counts between the fractions.

Large-scale microfluidic screening reveals regulators of α -amylase secretion

We assessed the potential of the platform by screening and sorting approximately 800 000 droplets, at a 8-fold coverage of the original library size (Table S1†), which among a smaller number of hits, identified guides targeting *PDI1*, an essential ER gene for which overexpression has previously been shown to increase protein secretion³¹ and *NMA2*, part of the NAD⁺ salvage pathway that protects from proteotoxicity by clearing misfolded proteins, and also acting as a chaperone.^{32,33} Encouraged by the results we scaled the screen to 2 million droplets to achieve a 20-fold library coverage, and pooled the results from multiple replicate screens during the statistical analysis. The combined results from multiple replicate rounds of screening identified 11 218 guides from the activation screens (Fig. 1F, Table S1†) and 20 561 guides from the repression screens (Fig. 1G, Table S1†) at sufficient coverage to meet the filtering threshold for the statistical analysis; this represents a 27% and 50% coverage of the total guide pool, resulting in a 83% and 97% representation of all genes with 1.6 and 2.9 guides per gene on average for the activation screens or the repression screens, respectively. The analysis identified 311 guides, out of these remaining guides, mapping to 306 individual genes as enriched [$\log FC > 3$, $FDR < 0.05$] for the activation screens, and a smaller set of 34 guides, mapping to 34 genes as enriched [$\log FC > 3$, $FDR < 0.05$] for the repression screens (Table S1†).



Network analysis identifies an interplay of processes affecting α -amylase secretion

To identify cellular processes and pathways that affect α -amylase secretion, we subjected a wider subset [$\log FC > 3$, $FDR < 0.3$] of enriched genes to GO enrichment analysis. Cellular pathways of protein maturation and secretion were

enriched (Fig. 2A). These included higher-level terms relating to vesicles, the Golgi apparatus, the periplasmic space, and cell surface for the activation screens. In contrast, for the repression screens (Fig. 2B), the majority of enriched GO terms for cellular components were not directly involved in protein secretion, with the signal recognition particle and the biological process of autophagy (Table S2, Fig. S1†) as exceptions. The screens also

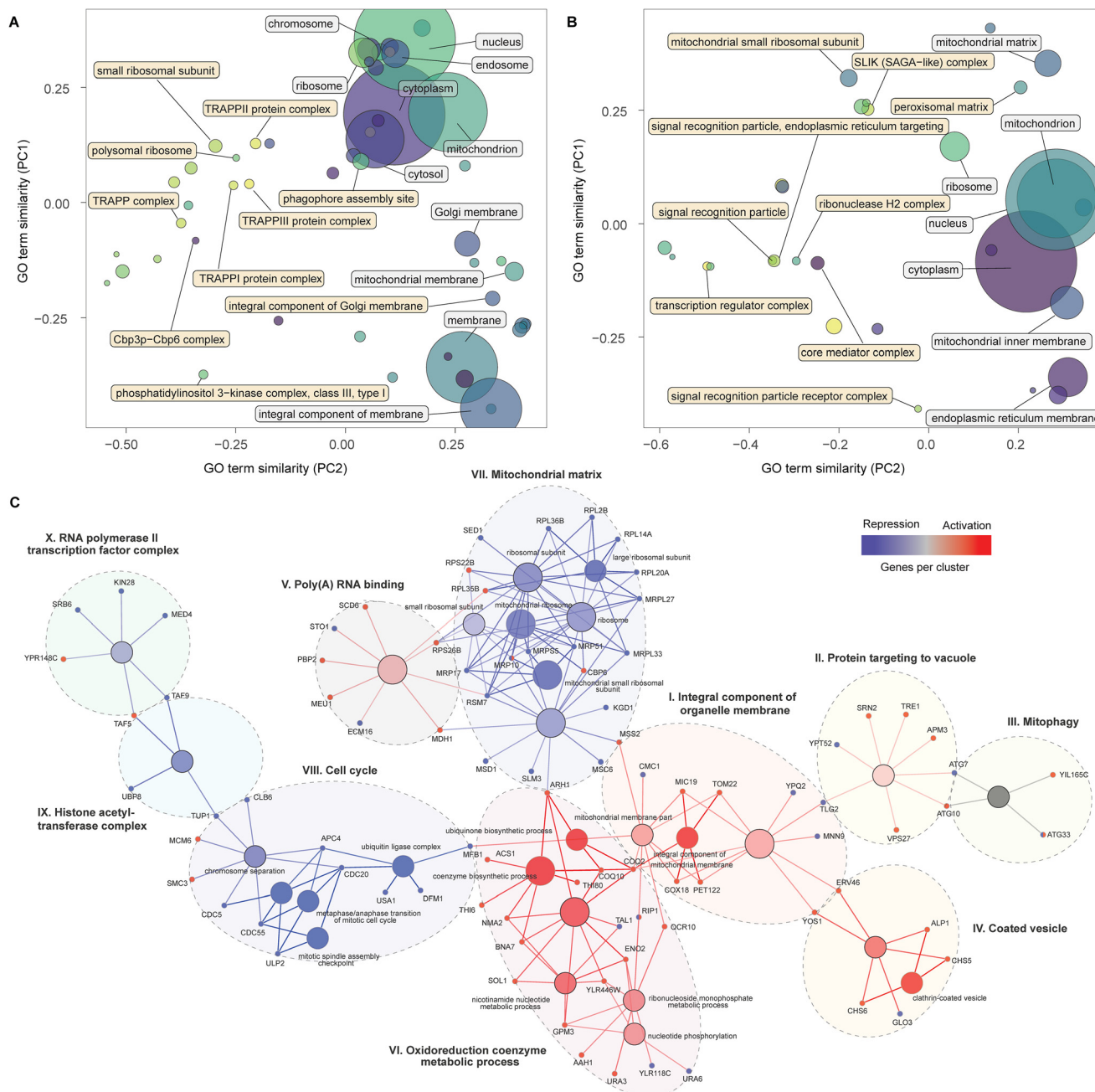


Fig. 2 Cellular components, nodes and genes modulating α -amylase secretion. PCA similarity of significantly enriched GO terms for cellular components for the activation (A) and repression screens (B), with terms identified by standard (white) and topology-based (yellow) analysis. The size of the circles represents the number of genes contained in the GO term. Groups of nodes and genes identified as significantly enriched by functional network analysis (C) combining the results from the activation and repression screens. Gene colours, blue (repression) and red (activation), identify the screens for which enrichment was detected, while nodes are coloured in a gradient based on the number of genes from each screen, from red (primarily repression), to blue (primarily activation) and grey (equal). Each cluster is identified by a roman numeral and the assigned name.



identified terms related to the cell cycle, gene regulation and translation, energy supply, biosynthetic processes for sugars, amino acids, and nucleosides as enriched for the activation libraries (Table S2, Fig. S1†). For the repression screens, enriched GO terms included chromosome separation and mitotic division, mitochondrial gene expression and translation (Table S2, Fig. S1†).

Next, we combined a smaller subset of genes (see Methods for details) from the activation and repression screens into a single network, displaying enriched pathways and GO-terms with a FDR < 0.05, in an attempt to provide an integrated map of genes and processes that increase α -amylase secretion. This analysis identified ten clusters (Fig. 2C), six of which were primarily composed of genes from the activation screens (I to VI). Cluster (I) contains genes encoding proteins in the inner membrane of mitochondria, including *COX18* and *PET122*, for which the null mutants are deficient in respiratory growth,^{34,35} and *TOM22*, which is central to protein import into mitochondria during the metabolic switch from fermentative to respiratory growth.^{36,37} Cluster (II) connects *via* *TLG2*, a gene encoding a t-SNARE protein involved in endosome to Golgi trafficking and the Cvt (cytoplasm-to-vacuole targeting) pathway³⁸ together with *YPT52* that enables localization of the CORVET complex to endosomes.³⁹ Other genes in this cluster, including *VPS27*, *TRE1*, and *SRN2*, are involved in the sorting of ubiquitinated proteins for degradation and autophagy (*ATG7*, *ATG10*, including *ATG33* from the mitophagy cluster (III)). The coated vesicle cluster (IV) contains genes linked to COPII transport from the ER to the Golgi, including *YOS1* and *ERV46*, and the two genes in the exomer complex (*CHS5* and *CHS6*) that deliver cargo to the plasma membrane. The poly(A) RNA binding cluster (V) includes *MDH1* part of the malate-aspartate shuttle that exchanges NADH between the cytosol and mitochondria. Overexpression of *MDH1* in *Pichia pastoris* resulted in a 40% increase in recombinant protein production.⁴⁰ This cluster also includes *SCD6* which represses translation in budding yeast,⁴¹ the ortholog of this gene (*tral*) in *Drosophila* is a positive regulator of protein secretion.⁴² The remaining cluster (VI) contains genes associated with ubiquinone biosynthesis (*COQ10*, *COQ2*, *ARH1*), as well as *BNA7* and *NMA2* linked to biosynthesis or recovery of NADH and other genes involved in nucleotide-related processes.

The largest cluster of genes from the repression screens encode cytoplasmic and mitochondrial ribosomal subunits (VII). This may be linked mechanistically to another cluster (VIII) containing genes relating to an arrest of the cell cycle. The cell cycle can be indirectly stalled in the G1 phase due to a lack of synthesis of new ribosomes⁴³ and directly arrested during multiple mitotic checkpoints *via* cell cycle division (CDC) genes,⁴⁴ several of which are found in this group. Considering that around 30% of the cells energy is allocated towards general protein synthesis during the cell cycle,⁴⁵ a reduction in growth would instead enable shunting of these resources towards synthesis of recombinant proteins. The

two remaining clusters (IX–X) included genes that encode components of the histone acetyltransferase complex and transcription factors for RNA polymerase II, highlighting the importance of transcriptional processes. *TUP1* from the first cluster is a general repressor of transcription,⁴⁶ while *TAF5* and *TAF9*, in the interface between the two clusters, are TATA-binding and involved in the initiation of transcription, around 60% of all yeast genes are dependent on *TAF9*, the highest number of dependent genes for all 13 TAF proteins.⁴⁷

Potential role for non-coding RNAs in modulation of α -amylase secretion

The activation and repression screens identified 71 and 8 SUTs (stable unannotated transcripts) or CUTs (cryptic unstable transcripts) respectively, as enriched. SUTs generate stable transcripts that are thought to interact with other transcripts, while CUTs are unstable and quickly degrade after transcription.⁴⁸ Transcripts of both classes primarily mediate effects on gene expression in *cis*.⁴⁹ An enrichment analysis (see Methods) of genes neighbouring the SUT or CUT (1 kb interval centred on the targeting guide) identified enriched GO terms for the set from the activation screens but not the repression screens. The most overrepresented terms, related to cellular components, contained genes from vacuolar, endosomal and Golgi related processes (Table S2†). Manual inspection of 17 genes found within the top 5 enriched groups of cellular components (Fig. S2†) based on the analysis, out of 102 identified within the range, showed that 15 SUTs or CUTs were situated on the opposite strand, four in a divergent position potentially acting as bidirectional promoters, resulting in gene activation, while eleven were either directly or in a few cases, indirectly through guide interference, overlapping, such that the CRISPRi likely resulted in a simultaneous decrease in the transcription of their antisense gene. Among these, SUT428 overlapping *OPT2* has previously been investigated in detail,⁵⁰ expression of this SUT resulted in a corresponding reduction in transcription of the overlapping gene. The two remaining SUTs or CUTs were found on the same strand, one directly downstream of the identified gene, and the other, CUT727 directly upstream of *BET3*, a core component of the TRAPP complexes. CUTs found upstream, overlapping the 5' transcriptional start site, may act in a regulatory role.⁵¹ Manual inspection of the small number of enriched SUTs or CUTs from the repression screen (examples in Fig. S3†), showed that CUT361 was situated directly upstream of *ATG19*, an autophagy gene involved in the Cvt pathway. SUT433 and CUT437 were notably situated in tandem, spaced approximately 1 kb apart with SUT433 situated directly upstream of *FLO9*, a flocculation gene, while CUT437 further upstream is in a potential bidirectional promoter position with *GDH3*, involved in glutamate biosynthesis, which has been identified as a target for increased protein secretion both in *S. cerevisiae*⁵² and *P. pastoris*.^{40,53} *FLO11*, another flocculation gene, which was not identified in this screen, has previously been shown to be regulated by two adjacent ncRNA's.⁵⁴



Validation of over-secretion genes quantified in the screens

We selected 16 genes identified as significant hits during the screening process for follow-up experimental validation of amylase over-secretion using alternative methods for overexpression or repression. Ten gene hits from the activation screens were validated *via* plasmid-based overexpression of the native gene, while six candidates from the repression screens were evaluated *via* gene deletion in both alleles of the diploid strain. The units of secreted α -amylase and the cell density (OD 600) were measured after 4, 24, and 48 hours of growth (Table S3†) and compared to the baseline expression (see Methods). Overexpression of all genes with the exception of TOM22 resulted in an increase of 35–60% in total α -amylase secretion after 4 hours of growth. This trend continued for *ENO2*, *NMA2*, *PRY2*, *SUT074* and *TFG2* with 20–40% increases in total α -amylase secretion after 24 and 48 hours (Fig. 3A). Secretion of α -amylase on a per cell basis (see Methods) was significantly increased for *TOM22* in the late stationary phase after 48 hours, and after 24 and 48 hours in *BNA7*. Gene deletions (Fig. 3B), resulted in increased total α -amylase secretion for *HDA2* (included as a positive control), *MNT2*, *TPO2* and *INP51* after 4 hours, while for *INP51* a significant decrease was seen after 4 hours, this decrease was not visible after 24 hours and was found to be significantly increased after 48 hours. An increase in

secretion of α -amylase on a per cell basis was seen in *INP51* under all time points, in *TLG2* after 4 hours and in *HDA2* after 4 and 24 hours. All over-expressed or deleted genes, with the exception of *TRS20* and *YDR262W* (Fig. S4†), resulted in increased α -amylase secretion under at least one time point.

Discussion

We applied large-scale CRISPR activation and repression libraries to perturb the expression of the majority of genes and ncRNAs in *S. cerevisiae*, to evaluate the effects on α -amylase secretion. The combination of large-scale CRISPRa/i screening and microfluidics with an alpha-amylase secretion assay adapted to droplet-encapsulated single cells, allowed us to identify 345 genes important for α -amylase secretion. This forms the to date most comprehensive genetic screening effort for α -amylase secretion in budding yeast. The results from activation and repression screens are complementary,⁵⁵ which allowed us to build gene networks and perform enrichment analysis to infer secretion-related cellular processes and molecular machineries from a substantially richer material compared to CRISPRa and CRISPRi alone. We found that increased expression of genes that were previously associated with cellular components linked to the secretory pathway resulted in increased

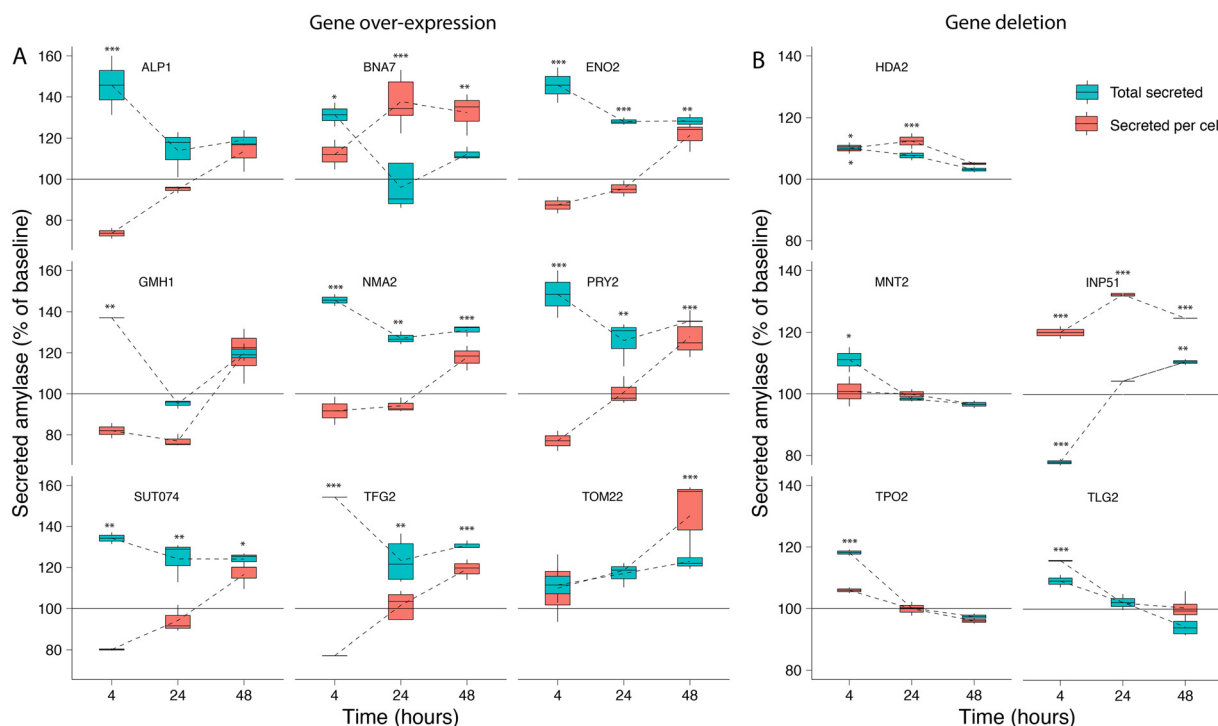
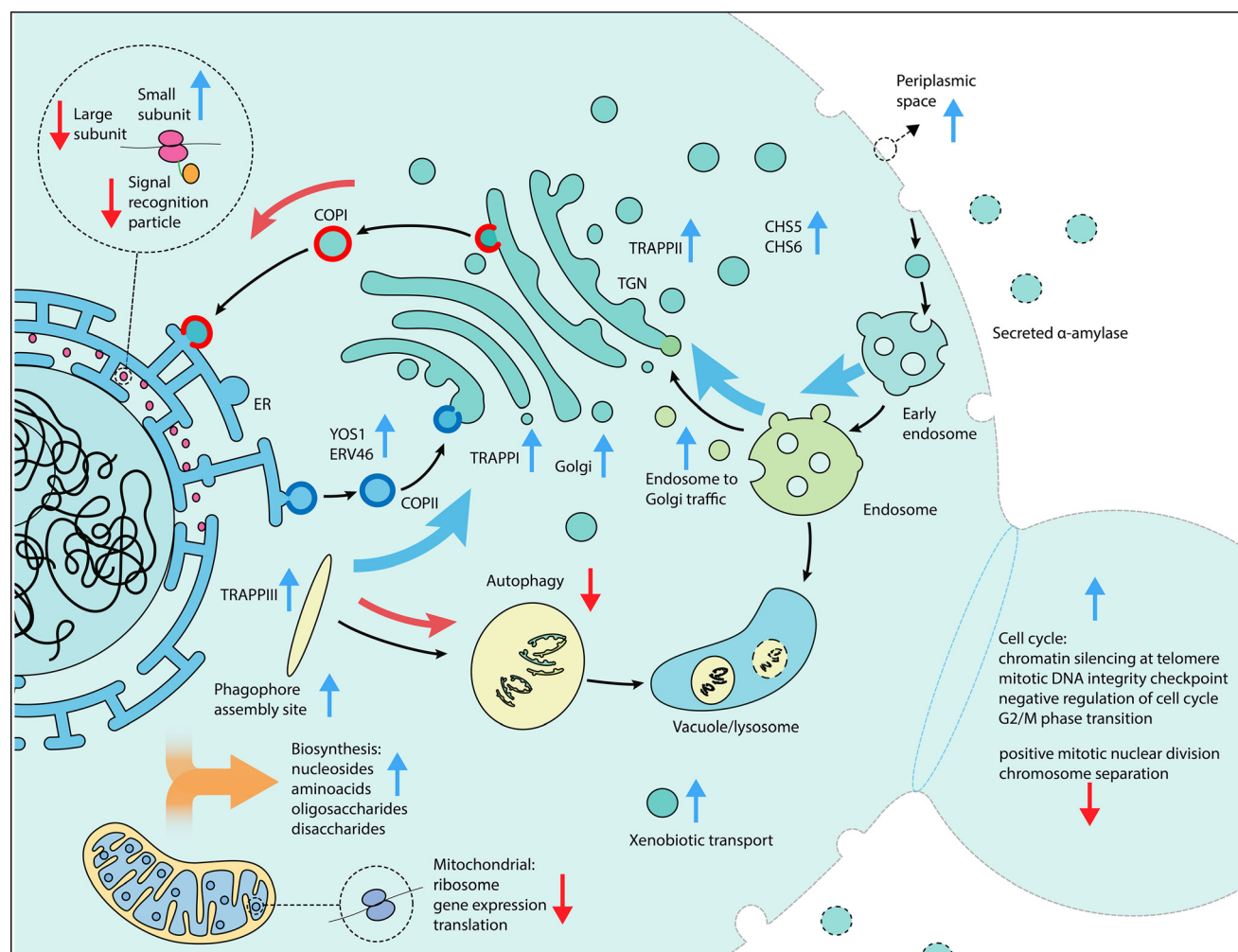


Fig. 3 Validation of amylase secretion. For selected genes identified during the activation and repression screens, the corresponding (A) over-expression and (B) deletion strains were generated and assayed for α -amylase secretion over time. The results are shown as total secreted α -amylase (blue) and secreted α -amylase per cell (red) for measurements after 4, 24, and 48 hours, relative to the baseline α -amylase secretion of the transformed empty vector for gene over-expression, or the background strain only for gene deletion. Error bars are one standard deviation from biological replicates ($n = 2-5$), values that are significantly different (ANOVA, $p < 0.05$) from the control (see Methods) within a time point are shown with asterisks corresponding to the p -value (***0.001, **0.01 or *0.05) from a Tukey HSD *post hoc* test.



Among the more unexpected results, we identified components of the signal recognition particle as enriched in the repression screen, suggesting that a reduction in co-

Among the more unexpected results, we identified components of the signal recognition particle as enriched in the repression screen, suggesting that a reduction in co-



This journal is © The Royal Society of Chemistry 2023

translational targeting of nascent protein to the ER is beneficial under some conditions. Balancing the overall metabolic demands of the cell is important. Heterologous expression of recombinant proteins can result in substantial reductions in growth rates due to the metabolic burden.⁶¹ The results from the activation screens show an enrichment of biosynthetic processes for sugars, including disaccharides and oligosaccharides, amino acids from the aspartate and lysine family, and nucleosides. Biosynthesis of amino acids and translation are the most energetically costly steps that are related to protein secretion.⁶² We further found an enrichment of processes related to NADPH regeneration and shunting which likely help to offset some of the associated energy demand (Table S2†). The activation screen identified negative regulation of the cell cycle *via* DNA-dependent DNA replication and mitotic DNA damage checkpoint and mitotic G2/M transition checkpoint as beneficial, in conjunction with suppression of chromosome separation, localization to the spindle body and positive mitotic nuclear division for the repression screen. Among these genes were *ENO2*, for which over-expression has previously been shown to result in G1 arrest.⁶³ Overexpression of this gene as part of our validation, resulted in increased amylase secretion (Fig. 3A). Together, these results suggest that an arrest of the cell cycle is beneficial for increased α -amylase trafficking and secretion. Additionally, the enrichment of genes in the small subunit of the mitochondrial ribosome and mitochondrial gene expression and translation in the repression screen also suggests that a temporary reduction of mitochondrial translation processes is beneficial for increased α -amylase secretion (Fig. 2B).

In our work, we also investigated the influence of long non-coding RNAs on α -amylase secretion. We identified enriched SUTs and CUTs in the vicinity of genes that encode for components in the interface of the late Golgi, endosome and vacuole. Of the manually inspected hits (Fig. S2†), the vast majority were situated on the opposite strand in a position that may result in interference of gene transcription.^{64,65} Notable examples in addition to the previously discussed SUT428 and CUT727 are CUT586 and CUT480. CUT586 overlaps *YPT32* on the opposite strand; *Ypt32* forms a GTPase with *Ypt31*, and regulates the exocytic pathway.⁶⁶ The TRAPP II complex, of which we found several components being enriched in our screen, in turn acts as the guanine exchange factor that activates *Ypt31/Ypt32*.⁶⁷ CUT480 overlaps with *ARF2*, a functionally interchangeable homologue of *ARF1*,⁶⁸ a GTPase with a closely related role to *Ypt31/Ypt32*, which further stimulates *SEC7* during sorting and localization to the late Golgi.⁶⁹ Further among the hits are CUT276, situated opposite of *VPS33*, one of the four core subunits of the HOPS and CORVET tethering complexes. The exact role and mechanisms of action of the identified SUTs and CUTs would need to be experimentally tested. Secondary effects, from guides within distance of the TSS, or overlapping the CDS of other genes, may affect the results.

We successfully used industrial production strains for genetic screens, instead of established laboratory strains.

Efficient protein secretion is a valuable trait for biotechnological protein production and the use of a robust industrial strain is beneficial for the interpretability of hits that might be relevant to commercial applications. Performing genetic screens in industrial strains, however, is challenging, since industrial strains are difficult to transform, with reported efficiencies 2 to 3 magnitudes lower compared to established laboratory strains.⁷⁰ This makes the use of these strains for genetic screens challenging since the assembly and transformation of the gRNA libraries has to be carried out at sufficient scale, estimated at around 20 times the library size,⁷¹ to retain the diversity of the original gRNA pool. Our work confirmed low transformation efficiency. Substantial optimization of strain transformation (see Methods section) allowed us to overcome this issue, which will be useful for future studies within industrially relevant strains. The *in vitro* validation of a subset of enriched genes from the secretion screens supported our findings from the pooled genetic screens (Fig. 3). In summary, over-expression of genes (*ENO2*, *NMA2*, *PRY2*, *SUT074*, *TFG2*), or gene deletion (*INP51*) resulted in around 20–30% increases in total α -amylase secretion, sustained over 24 and 48 hours. With higher increases (20–40%) on a per cell basis (*BNA7*, *TOM22*, *INP51*). The majority of the validated genes have, to our knowledge, not previously been identified as targets for increased α -amylase secretion.

We anticipate that this study can serve as a blueprint for future investigations into the effects of gene expression perturbation as well as gene editing screens on protein secretion. The use of a CRISPR/dCas9 system with guides acting as barcodes, combined with high throughput microfluidics, substantially increases the speed and scale of discovery compared to previous studies.^{7–9} Based on the time requirements of our large-scale screens with tens of thousands of perturbations, which were typically performed over a week, with one day of screening. We believe that smaller screens using targeted libraries, would be able to generate results of similar or higher quality in a matter of days, including transformation, screening and sequencing. This would enable iterative testing and optimization of new strains for improved protein secretion. Future developments could also include the use of CRISPR-based precision genome editing, such as MAGESTIC,⁷² which would allow for probing of natural variants in relevant strains with desirable traits.

Data availability

Raw and processed sequencing data are available at Gene Expression Omnibus (GEO) under accession number GSE221655.

Author contributions

S. A. J. and W. W. conceived the study. S. A. J. designed and performed experiments, analysed data and wrote the manuscript. A. F. designed and generated secretion strains.



T. D. and C. R. assisted with validation experiments and expertise. W. W. designed and K. R. R. and J. D. S. generated gRNA libraries and plasmids. A. P. designed the microfluidic devices and together with C. A. M. provided expertise and supervision. T. D., C. J., J. D. S., W. W., K. R. R., C. A. M., G. P., D. S., D. C., T. D., A. F. and L. M. S. provided valuable advice and or revised the manuscript. L. M. S. supervised the study.

Conflicts of interest

Authors T. D., G. P., D. C., T. D., C. R., A. F., and J. D. S. (via Recombia Biosciences) are or were full-time employees of Lesaffre Group who exclusively funded the study. L. M. S. is co-founder and shareholder of Sophia Genetics, co-founder and board member of LevitasBio and Recombia Biosciences.

Acknowledgements

We are grateful to all members of the EMBL Genomics core facility for their support and help with sequencing related questions, especially Vladimir Benes.

Notes and references

- P. Novick and R. Schekman, *Proc. Natl. Acad. Sci. U. S. A.*, 1979, **76**, 1858–1862.
- K. Takeshige, M. Baba, S. Tsuboi, T. Noda and Y. Ohsumi, *J. Cell Biol.*, 1992, **119**, 301–311.
- K.-K. Hong and J. Nielsen, *Cell. Mol. Life Sci.*, 2012, **69**, 2671–2690.
- Protein Expression Market worth \$4.1 billion by 2025, <https://www.marketsandmarkets.com/PressReleases/protein-expression.asp>, (accessed July 8, 2021).
- C.-C. Kuo, A. W. T. Chiang, H. M. Baghdassarian and N. E. Lewis, *Cell Syst.*, 2021, **12**, 873–884.
- J. Tyedmers, A. Mogk and B. Bukau, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 777–788.
- S. L. Sjöström, Y. Bai, M. Huang, Z. Liu, J. Nielsen, H. N. Joensson and H. A. Svahn, *Lab Chip*, 2013, **14**, 806–813.
- M. Huang, Y. Bai, S. L. Sjöström, B. M. Hallström, Z. Liu, D. Petranovic, M. Uhlén, H. N. Joensson, H. Andersson-Svahn and J. Nielsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E4689–E4696.
- G. Wang, S. M. Björk, M. Huang, Q. Liu, K. Campbell, J. Nielsen, H. N. Joensson and D. Petranovic, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 9324–9332.
- J. D. Smith, S. Suresh, U. Schlecht, M. Wu, O. Wagih, G. Peltz, R. W. Davis, L. M. Steinmetz, L. Parts and R. P. S. Onge, *Genome Biol.*, 2016, **17**, 1–16.
- A. Chavez, J. Scheiman, S. Vora, B. W. Pruitt, M. Tuttle, E. P. R. Iyer, S. Lin, S. Kiani, C. D. Guzman, D. J. Wiegand, D. Ter-Ovanesyan, J. L. Braff, N. Davidsohn, B. E. Housden, N. Perrimon, R. Weiss, J. Aach, J. J. Collins and G. M. Church, *Nat. Methods*, 2015, **12**, 326–328.
- A. A. Dominguez, W. A. Lim and L. S. Qi, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 5–15.
- Z. Xu, W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, J. Camblong, E. Guffanti, F. Stutz, W. Huber and L. M. Steinmetz, *Nature*, 2009, **457**, 1033–1037.
- J. Listgarten, M. Weinstein, B. P. Kleinstiver, A. A. Sousa, J. K. Joung, J. Crawford, K. Gao, L. Hoang, M. Elibol, J. G. Doench and N. Fusi, *Nat. Biomed. Eng.*, 2018, **2**, 38–47.
- R. S. Sikorski and P. Hieter, *Genetics*, 1989, **122**, 19–27.
- L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman and L. S. Qi, *Cell*, 2013, **154**, 442–451.
- L. Benatui, J. M. Perez, J. Belk and C.-M. Hsieh, *Protein Eng., Des. Sel.*, 2010, **23**, 155–159.
- C. Chaipan, A. Pryszlak, H. Dean, P. Poignard, V. Benes, A. D. Griffiths and C. A. Merten, *Cell Chem. Biol.*, 2017, **24**, 751–757.
- C. Girardot, J. Scholtalbers, S. Sauer, S.-Y. Su and E. E. M. Furlong, *BMC Bioinf.*, 2016, **17**, 1–6.
- H. Li and R. Durbin, *Bioinformatics*, 2009, **25**, 1754–1760.
- M. D. Robinson, D. J. McCarthy and G. K. Smyth, *Bioinformatics*, 2010, **26**, 139–140.
- D. J. McCarthy, Y. Chen and G. K. Smyth, *Nucleic Acids Res.*, 2012, **40**, 4288–4297.
- Y. Benjamini and Y. Hochberg, *J. R. Stat. Soc. Series B Stat. Methodol.*, 1995, **57**, 289–300.
- L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo and H. Peterson, *F1000Research*, 2020, **9**, 1–27.
- A. Alexa and J. Rahnenfuhrer, *topGO: Enrichment Analysis for Gene Ontology. R package version 2.42.0.*, 2020.
- A. Brionne, A. Juanchich and C. Hennequet-Antier, *BioData Min.*, 2019, **12**, 1–16.
- G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski and J. Galon, *Bioinformatics*, 2009, **25**, 1091–1093.
- G. Bindea, J. Galon and B. Mlecnik, *Bioinformatics*, 2013, **29**, 661–663.
- L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin and W. A. Lim, *Cell*, 2013, **152**, 1173–1183.
- M. M. Demeke, H. Dietz, Y. Li, M. R. Foulquié-Moreno, S. Mutturi, S. Deprez, T. D. Abt, B. M. Bonini, G. Liden, F. Dumortier, A. Verplaetse, E. Boles and J. M. Thevelein, *Biotechnol. Biofuels*, 2013, **6**, 1–24.
- A. S. Robinson, V. Hines and K. D. Wittrup, *Bio/Technology*, 1994, **12**, 381–384.
- M. Pinkerton, A. Ruetenik, V. Bazylanska, E. Nyvltova and A. Barrientos, *Hum. Mol. Genet.*, 2021, **30**, 672–686.
- A. Ocampo, J. Liu and A. Barrientos, *Hum. Mol. Genet.*, 2013, **22**, 1699–1708.
- Å. Valadi, K. Granath, L. Gustafsson and L. Adler, *J. Biol. Chem.*, 2004, **279**, 39677–39685.
- E. Espinoza-Simón, N. Chiquete-Félix, L. Morales-García, U. Pedroza-Dávila, X. Pérez-Martínez, D. Araiza-Olivera, F. Torres-Quiroz and S. Uribe-Carvajal, *Fungal Biol.*, 2020, **124**, 15–23.
- C. Gerbeth, O. Schmidt, S. Rao, A. B. Harbauer, D. Mikropoulou, M. Opalińska, B. Guiard, N. Pfanner and C. Meisinger, *Cell Metab.*, 2013, **18**, 578–587.



- 37 S. van Wilpe, M. T. Ryan, K. Hill, A. C. Maarse, C. Meisinger, J. Brix, P. J. T. Dekker, M. Moczko, R. Wagner, M. Meijer, B. Guiard, A. Hönlanger and N. Pfanner, *Nature*, 1999, **401**, 485–489.
- 38 H. Abeliovich, T. Darsow and S. D. Emr, *EMBO J.*, 1999, **18**, 6005–6016.
- 39 M. Cabrera, H. Arlt, N. Epp, J. Lachmann, J. Griffith, A. Perz, F. Reggiori and C. Ungermann, *J. Biol. Chem.*, 2013, **288**, 5166–5175.
- 40 J. Nocon, M. G. Steiger, M. Pfeffer, S. B. Sohn, T. Y. Kim, M. Maurer, H. Rußmayer, S. Pflügl, M. Ask, C. Haberhauer-Troyer, K. Ortmayr, S. Hann, G. Koellensperger, B. Gasser, S. Y. Lee and D. Mattanovich, *Metab. Eng.*, 2014, **24**, 129–138.
- 41 D. Roy and P. I. Rajyaguru, *Wiley Interdiscip. Rev.: RNA*, 2018, **9**, 1–17.
- 42 J. E. Wilhelm, M. Buszczak and S. Sayles, *Dev. Cell*, 2005, **9**, 675–685.
- 43 K. A. Bernstein and S. J. Baserga, *Mol. Biol. Cell*, 2004, **15**, 5038–5046.
- 44 T. A. Weinert and L. H. Hartwell, *Genetics*, 1993, **134**, 63–80.
- 45 F. Buttgereit and M. D. Brand, *Biochem. J.*, 1995, **312**, 163–167.
- 46 R. L. Smith and A. D. Johnson, *Trends Biochem. Sci.*, 2000, **25**, 325–330.
- 47 W. Shen, S. R. Bhaumik, H. C. Causton, I. Simon, X. Zhu, E. G. Jennings, T. Wang, R. A. Young and M. R. Green, *EMBO J.*, 2003, **22**, 3395–3402.
- 48 D. Wilkinson, L. Váchová, O. Hlaváček, J. Maršíková, G. D. Gilfillan and Z. Palková, *Oxid. Med. Cell. Longevity*, 2018, **2018**, 1–12.
- 49 S. Parker, M. G. Fraczek, J. Wu, S. Shamsah, A. Manousaki, K. Dungrattanalert, R. A. de Almeida, E. Invernizzi, T. Burgis, W. Omara, S. Griffiths-Jones, D. Delneri and R. T. O'Keefe, *PLoS Genet.*, 2018, **14**, 1–33.
- 50 M. C. Marvin, S. Clauder-Münster, S. C. Walker, A. Sarkeshik, J. R. Yates, L. M. Steinmetz and D. R. Engelke, *RNA*, 2011, **17**, 1441–1450.
- 51 H. Neil, C. Malabat, Y. d'Aubenton-Carafa, Z. Xu, L. M. Steinmetz and A. Jacquier, *Nature*, 2009, **457**, 1038–1042.
- 52 M. Huang, J. Bao, B. M. Hallström, D. Petranovic and J. Nielsen, *Nat. Commun.*, 2017, **8**, 1–12.
- 53 M. Carnicer, A. ten Pierick, J. van Dam, J. J. Heijnen, J. Albiol, W. van Gulik and P. Ferrer, *Microb. Cell Fact.*, 2012, **11**, 1–10.
- 54 S. L. Bumgarner, R. D. Dowell, P. Grisafi, D. K. Gifford and G. R. Fink, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 18321–18326.
- 55 L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann and J. S. Weissman, *Cell*, 2014, **159**, 647–661.
- 56 J. J. Kim, Z. Lipatova and N. Segev, *Front. Cell Dev. Biol.*, 2016, **4**, 1–11.
- 57 C. K. Yip, J. Berscheminski and T. Walz, *Nat. Struct. Mol. Biol.*, 2010, **17**, 1298–1304.
- 58 S. Zhao, C. M. Li, X. M. Luo, G. K. Y. Siu, W. J. Gan, L. Zhang, W. K. K. Wu, H. C. Chan and S. Yu, *Sci. Rep.*, 2017, **7**, 1–12.
- 59 W. van Leeuwen, F. van der Krift and C. Rabouille, *J. Cell Biol.*, 2018, **217**, 2261–2271.
- 60 C. Anton-Plagaro, N. Sanchez, R. Valle, J. M. Mulet, M. C. Duncan and C. Roncero, *FASEB J.*, 2021, **35**, 1–26.
- 61 R. Gonzalez, B. A. Andrews, J. Molitor and J. A. Asenjo, *Biotechnol. Bioeng.*, 2003, **82**, 152–169.
- 62 T. Klein, J. Niklas and E. Heinzle, *J. Ind. Microbiol. Biotechnol.*, 2015, **42**, 453–464.
- 63 W. Niu, Z. Li, W. Zhan, V. R. Iyer and E. M. Marcotte, *PLoS Genet.*, 2008, **4**, 1–15.
- 64 Z. Xu, W. Wei, J. Gagneur, S. Clauder-Münster, M. Smolik, W. Huber and L. M. Steinmetz, *Mol. Syst. Biol.*, 2011, **7**, 468–468.
- 65 A. Nevers, A. Doyen, C. Malabat, B. Néron, T. Kergrohen, A. Jacquier and G. Badis, *Nucleic Acids Res.*, 2018, **46**, 6009–6025.
- 66 Z. Lipatova and N. Segev, *FEBS Lett.*, 2019, **593**, 2488–2500.
- 67 C. Mi, L. Zhang, G. Huang, G. Shao, F. Yang, X. You, M.-Q. Dong, S. Sun and S.-F. Sui, *Sci. Adv.*, 2022, **8**, 1–15.
- 68 T. Stearns, R. A. Kahn, D. Botstein and M. A. Hoyt, *Mol. Cell Biol.*, 1990, **10**, 6690–6699.
- 69 C. M. McDonold and J. C. Fromme, *Dev. Cell*, 2014, **30**, 759–767.
- 70 T. Ogata, Y. Okumura, M. Tadenuma and G. Tamura, *J. Gen. Appl. Microbiol.*, 1993, **39**, 285–294.
- 71 T. Wang, C. Guan, J. Guo, B. Liu, Y. Wu, Z. Xie, C. Zhang and X.-H. Xing, *Nat. Commun.*, 2018, **9**, 1–15.
- 72 K. R. Roy, J. D. Smith, S. C. Vonesch, G. Lin, C. S. Tu, A. R. Lederer, A. Chu, S. Suresh, M. Nguyen, J. Horecka, A. Tripathi, W. T. Burnett, M. A. Morgan, J. Schulz, K. M. Orsley, W. Wei, R. S. Aiyar, R. W. Davis, V. A. Bankaitis, J. E. Haber, M. L. Salit, R. P. S. Onge and L. M. Steinmetz, *Nat. Biotechnol.*, 2018, **36**, 512–520.

