nature catalysis

REVIEW ARTICLE

https://doi.org/10.1038/s41929-022-00836-w

Check for updates

Towards next-generation cell factories by rational genome-scale engineering

Suzan Yilmaz[1], Akos Nyerges[2], John van der Oost[1], George M. Church[2,3,4] and Nico J. Claassens[1]✉

Metabolic engineering holds the promise to transform the chemical industry and to support the transition into a circular bio-economy, by engineering cellular biocatalysts that efficiently convert sustainable substrates into desired products. However, despite decades of research, the potential of metabolic engineering has only been realized to a limited extent at the industrial level. To further realize its potential, it is essential to optimize the synthetic and native metabolic networks of cell factories at a system and genome-wide level. Here we discuss the tools and strategies enabling system-wide (semi-) rational engineering. Recent advances in genome-editing technologies enable directed genome-wide engineering in a growing number of relevant microorganisms. Such system-wide engineering can benefit from machine learning and other in silico design methods, and it needs to be integrated with efficient screening or selection approaches. These approaches are expected to realize the promise of next-generation cell factories for efficient, sustainable production of a wide range of products.

Microbial cell factories have long been expected to become a key platform for the bioconversion of sustainable substrates into valuable products. The field of metabolic engineering emerged in the early 1990s[1] with the promising potential to revolutionize production processes in the chemical, fuel, pharma and food industries. In the past three decades, substantial technological advancements in DNA synthesis and assembly technologies[2], next-generation DNA sequencing[3], as well as genome engineering techniques[4–6] have greatly expanded the toolkit available to metabolic engineers. A vast amount of fundamental and applied research has been performed in the field, including the engineering of substrate utilization pathways for a range of cheap, sustainable feedstocks as well as of biosynthetic pathways for a broad spectrum of valuable natural and non-natural products[7,8]. These efforts have led to an impressive collection of proof-of-principle pathways and engineered strains[9]. Despite these success stories, however, bioproduction of commodity chemicals by metabolically engineered strains has only been realized so far to a limited extent at an industrial scale, primarily due to the unsatisfactory performance of many bioproduction processes that limits their economic feasibility[10].

A recent review of the industrial production of bio-based commodity chemicals and fuels identified approximately 30 different bio-based chemicals that are produced or are planned to be produced at commercial scale, which is a tiny number of compounds relative to the vast product portfolio of the petrochemical industry[9]. Similarly, the market share of bio-based chemical production (<US$10 billion in 2020)[11] is dwarfed by classical chemistry-based production (US$5.7 trillion per year)[12].

The current bio-based market is primarily based on glucose as a substrate, which raises important sustainability concerns for future scale-up, as the production of glucose depends heavily on scarce agricultural resources and competes with food production[11,13]. So far, there are only a handful of bio-based industrial processes that are fed with more sustainable feedstocks such as lignocellulosic (residual) biomass, syngas (H_2/CO waste gas stream from some industries) or feedstocks generated with renewable electrical power, such as H_2 or reduced one-carbon molecules (for example, methanol and formate)[14–16].

To truly realize the promise of bio-based production, major challenges need to be tackled to allow the development of sustainable and economically viable bioprocesses. At the strain level, this includes strategies to realize efficient assimilation of sustainable and cheap substrates, as well as to increase the performance in terms of production titres, rates and yields. In addition, undesired effects related to scale-up from laboratory cultivation to large bioreactors need to be evaluated and addressed.

A major complication in generating well-performing next-generation cell factories is the complexity of cellular metabolic networks. Hence, a desired strategy towards improving production would be the simultaneous optimization of multiple metabolic bottlenecks and fluxes at a system level[10,17–19]. A valuable approach for system-level engineering is adaptive laboratory evolution (ALE)[20], for example to optimize substrate utilization, growth-coupled production and product tolerance. However, because ALE is typically based on random mutagenesis, the number of possible mutations is immense and only a small fraction of those will be beneficial. Hence, a powerful selection strategy is needed (that is, typically, selection for enhanced growth rates) and, as a result, ALE can often not be harnessed to optimize growth-competitive production. Even when it is possible to couple production to growth, the emergence of escape mutants that uncouple this connection may be a problem[21]. Moreover, phenotypic improvements of metabolic networks are often multifactorial problems requiring multiple epistatic mutations, which have a lower probability of occurring together and sometimes may even be unable to emerge due to evolutionary paths with local optima.

A promising alternative or complementary approach to ALE is provided by (semi)-rational, system-level engineering. In recent years, several genetic engineering techniques have been developed that allow for highly effective, directed, system-level engineering.

[1]Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands. [2]Department of Genetics, Harvard Medical School, Boston, MA, USA. [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [4]Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA, USA. ✉e-mail: nico.claassens@wur.nl

Advanced tools have been established, especially for the genetic engineering of the bacterial and eukaryotic model production organisms, respectively *Escherichia coli* and *Saccharomyces cerevisiae*, but also in several other less-studied hosts. The application of targeted system-level engineering can lead to many combinations of mutations to be tested; hence, these approaches require effective selection methods or screening strategies to identify variants with improved phenotypes. In this Review we discuss advances and emerging possibilities in system-level targeted metabolic engineering, including in silico target design strategies, high-throughput genome engineering techniques, as well as suitable screening and selection approaches.

Target prediction for systems metabolic engineering

Genetic targets for editing should be chosen carefully to limit the size of mutant libraries and make optimal use of the available screening capacity. The main difficulty in targeted metabolic engineering approaches is the requirement for prior knowledge about the metabolic network of the production organism as well as regulation of this network. However, increased understanding of cellular metabolic networks and in silico tools, such as metabolic modelling or machine learning, can support rational target selection.

Types of target gene

Metabolic engineering projects generally start with a pathway design for a product or substrate of interest, often designed or analysed with the help of in silico tools[22]. Subsequently, native and/or heterologous genes encoding the enzymes in the designed pathway are overexpressed either from plasmids or from the host chromosome. The latter approach is often preferred, as genomically integrated genes and modifications are more stable and avoid the use of antibiotics for plasmid selection. However, the mere introduction and overexpression of the pathway genes commonly results only in proof-of-principle, low-level productivities. Also, for sustainable substrate utilization pathways, full growth on these substrates can rarely be realized after initial pathway introduction attempts. Improving cell factories towards industrial performance usually requires extensive and iterative optimization of the flux in the pathway of interest and throughout the native metabolic network. Production performance can be optimized by preventing by-product formation and by adjusting metabolic bottlenecks, which involves targeting of coding and/or non-coding DNA regions so as to tune the performance of some key players, such as enzymes, transporters and transcriptional regulators. Targets can be modified via a range of intervention strategies, such as expression-level tuning, gene knockouts, as well as protein engineering strategies, in which (some) amino-acid residues are changed.

Generally, genes encoding the enzymes directly involved in the pathway of interest (for example, substrate assimilation or product biosynthesis) provide a straightforward starting point for targeting. In addition, native host genes may be identified as suitable targets by rationally inspecting the host metabolic network for pathways or regulatory mechanisms that can (directly or indirectly) impact strain performance. For example, metabolites can be converted by multiple, competing enzymes, often towards biosynthetic routes required for cellular growth or towards undesired by-products. The competing enzymes at branching points in a network are key targets for metabolic engineers. Traditionally, these targets are often knocked out in a time-consuming, iterative, trial-and-error process. In a more advanced system-level metabolic engineering approach, these targets are deleted simultaneously in different combinations. Alternatively, rather than knocking out these genes, more refined knockdown strategies allow for tuning enzyme levels and can potentially lead to an improved balance of production and growth. In addition to enzymes at branching points, targets include enzymes involved in biosynthesis or regeneration of cofactors that are important for cell factory performance, as well as transporters of relevant molecules (for example, substrates to be taken up from the medium or products to be excreted).

Another class of useful targets for system-level engineering comprises transcription factors. These regulators can control sets of genes at varying levels, ranging from specific regulators that control the expression of one or a few genes to global regulators that control the expression of dozens of genes. Early system-wide approaches to globally reprogram cells included the introduction of random mutations in subunits of the RNA polymerase complexes in *E. coli* or *S. cerevisiae*[23,24]. Such global regulators can change the expression patterns of many genes, but changes are hard to predict rationally, and require powerful screening or selection strategies. Still, targeting such regulators can be a useful approach, especially when it is poorly understood which specific genes should be targeted for a certain phenotype of interest, as is, for example, often the case for increasing host tolerance to toxic substrate compounds or products. Recent studies have targeted several dozens of transcriptional regulators in *E. coli* in parallel, trying to increase the tolerance to substances such as furfural, a toxic compound in lignocellulosic feedstocks, and styrene, a versatile but toxic monomer for bio-based plastics. In both studies, mutants of *E. coli* transcriptional regulators were identified that increase the tolerance level and hence increase productivities[25,26].

Despite the demonstrated successful examples of mutating transcriptional regulators, the complex regulatory effects of these targets may not necessarily lead to optimal outcomes. Mutations in (global) regulators may cause both beneficial and adverse alterations in expression at the same time, leading to unwanted trade-offs in cell factory optimization. Improved understanding of regulatory networks could help to identify promising, more specific regulators. However, metabolic models accounting for regulation have only started emerging for a limited number of model organisms, such as *E. coli* and *S. cerevisiae*[27,28].

In silico tools for target selection

The selection of targets is regularly done in a manual fashion based on metabolic and regulatory knowledge of the host organism, often including targets identified in previous engineering attempts. However, there has been a steady increase in the use of in silico modelling tools in the metabolic engineering field, which certainly may contribute to target selection.

The two most popular in silico modelling approaches for metabolic engineering are stoichiometric (or constraint-based) models and kinetic (or dynamic) models[29]. Stoichiometric models only require information on the reaction stoichiometry, and possibly some reaction constraints like upper or lower bounds of metabolite (in)fluxes, and reaction directionality[29,30]. A commonly applied type of stoichiometric modelling is flux balance analysis (FBA), in which linear programming is used to find flux distribution solutions within constraints for a given objective (for example, maximum growth or maximum productivity)[31]. FBA is generally performed using genome-scale metabolic models, which contain all stoichiometric reactions assumed to take place within a host based on enzymes annotated in the genome. As of now, these metabolic models have been constructed for many relevant microorganisms. To aid in target selection for metabolic engineering, some dedicated in silico tools based on FBA have been created. Generally, these tools predict sets of gene knockouts that are expected to increase the flux towards the desired product. Examples include OptKnock[32] or Minimal Cut Set (MCS) analysis[33]. Such tools can be used to limit the set of genetic knockout targets to be tested[34].

However, stoichiometric modelling generally does not consider factors like kinetics and regulation, so it often does not lead to the desired full performance. Furthermore, recent progress in the development of genetic tools allows for more subtle intervention

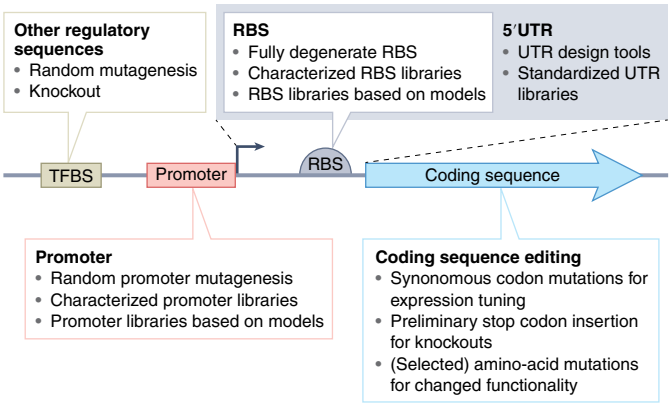

Fig. 1 | Target sites for gene (expression) editing and available modification strategies. Potential target sites for genetic modifications to modify expression levels include the transcription factor binding site (TFBS), the promoter, the 5′UTR, the RBS (only in prokaryotes) and the coding sequence (CDS). These different genomic sections can be adjusted through different modification strategies, each with specific levels of predictability and impact.

strategies than the creation of gene knockouts. For some microbial species, genetic tools are now available to modulate the expression strengths of multiple genes simultaneously, allowing for upregulation or downregulation of enzyme levels, which consequently may influence related fluxes. FBA-based tools can be used to some extent to predict such interventions. Ideally, however, models incorporating kinetics are used to identify potential targets for adjusting enzyme levels. A range of kinetic models have been proposed for metabolic engineering, but they mostly require reliable experimentally determined parameters of enzymatic kinetics[35,36]. The poor availability of such data is a major limitation in the application of these models.

To solve this problem, kinetic modelling approaches have been developed that do not rely on exact kinetic parameters. For example, ensemble modelling for robustness analysis (EMRA) uses a range of kinetic parameters to estimate which enzymatic reactions are crucial for the activity of an engineered pathway[37]. Such an EMRA framework has been applied to rationally predict some enzymes in the *E. coli* metabolic network that may need to be modulated to realize growth on methanol, via the non-native ribulose monophosphate (RuMP) pathway. Targeting of some of the enzymes predicted by this framework recently led to the long-awaited breakthrough of *E. coli* growing on the alternative substrate methanol as sole carbon and energy source via the RuMP cycle[38]. As suggested by the model, two reactions in glycolysis draining a key metabolite of the RuMP pathway (fructose 6-phosphate) were down-tuned by replacing them with slower enzymatic counterparts. However, these two interventions were insufficient to achieve full growth on methanol as the sole carbon and energy source. Only after several rounds of ALE was the full methylotrophic growth phenotype achieved; this included several more mutations, including mutations in the central carbon metabolism (for example, in the tricarboxylic acid cycle) that were not predicted by EMRA. Hence, to increase the success of rational metabolic engineering, additional complementary approaches besides metabolic models are important to help in predicting targets.

Other promising in silico tools include data-driven statistical and machine learning approaches, which generally use data from previous engineering iterations to inform the next round of engineering. One statistical type of method is the design of experiments (DoE), which designs a limited set of combinations to be tested in a first iteration, which can be effective in disentangling factors that are key to focus on in next iterations. A technical challenge of this approach is that only specific combinations of target variations should be generated, which is often not possible with the combinatorial genome engineering techniques, which generally vary the different targets independently. Still, the DoE approach has shown its power in the efficient optimization of heterologous pathways for which specific combinatorial libraries were cloned in vitro on plasmids and then tested in a microbial host[39–41].

Furthermore, machine learning algorithms can be used for target predictions based on data from previous iterations. Machine learning requires data for training, for example, productivities linked to mutant sequences of targets from previously engineered strains. Generally, machine learning will rely on an initial engineering iteration round with a wide set of selected targets and variants, which in next iterations can be narrowed down for further optimization. Only a few examples exist so far of the application of machine learning to metabolic engineering[42–44]. The potential of this approach for metabolic engineering is extensively reviewed elsewhere[45–47]. Although a potential drawback of machine learning is the requirement for extensive experimental data to train the model, it also holds promise for limiting the number of variants that need to be generated and screened in iterative rounds of system-wide engineering. Recently it was also demonstrated that machine learning can be used to address the lack of data for enzyme kinetic parameters, for example, by predicting enzyme turnover rates[48]. By using this approach, more accurate metabolic models could be generated for *S. cerevisiae*.

Genetic sequence targets and modification strategies

After the selection of target genes for modification, a next step is to identify regions of these genes to edit. The most frequent targets for editing include promoters, untranslated regions, ribosome binding sites (RBSs) or coding sequences (Fig. 1).

The promoter sequence is a key region with which to control expression levels, as it can regulate transcription initiation rates in both prokaryotic and eukaryotic cell factories. However, despite extensive studies, it is still poorly understood how natural promoter sequences control transcription levels. Only recently have some models been developed for predicting promoter strength in *E. coli*[49,50]. So far, promoter strength modulation in metabolic engineering is mostly performed using small libraries of well-known promoters with characterized strengths. However, a drawback of this approach is that the insertion of complete synthetic promoters in the genome requires larger sequence modifications (typically >30 bases), which are often challenging to realize with the available high-throughput genome-engineering techniques. An early example of genome-wide promoter replacement for metabolic engineering was demonstrated for the production of the dye compound indigo by *E. coli*[51]. Recently, several fundamental studies on promoter activities have shown that, by randomly mutating promoter regions, a wide, yet unpredictable range of expression strengths can be reached in, for example, *E. coli*[52]. Hence, making small random mutations in an existing promoter region can be an alternative method for system-wide promoter engineering, supported by tools such as the recombineering-based diversification tool DIvERGE[53], of which more details will be provided below. However, the trial-and-error nature of such a randomized promoter engineering approach requires efficient screening or selection of many variants.

Another key region for controlling protein production levels comprises the untranslated regions (UTRs) of messenger RNA (mRNA), especially the 5′UTR. This region plays crucial roles in the stability and degradation of mRNAs[54], and can typically cause variations in the enzyme levels as mRNA is translated into protein (for example, enzymes). In addition, in prokaryotes, the 5′UTR contains the RBS, which is responsible for recruiting ribosomes to initiate translation. For many bacterial species there is a fair

understanding of how RBS sequences determine translation initiation strength. The initiation rate is, amongst other factors, determined by base-pairing between six to eight bases of the RBS sequence in the mRNA and the 3′ end of the prokaryotic 16S rRNA, as well as by the accessibility of the RBS within a folded mRNA structure[55].

Some metabolic engineering efforts rely on libraries of previously characterized RBS sequences[56] or on large libraries of degenerate sequences covering the RBS and/or the first few codons of the coding sequence[57]. However, it must be noted that even short, degenerate RBS sequences of six bases already lead to a large library of $4^6 = 4{,}096$ variants for one gene. To reduce library sizes, computer algorithms such as RBS Calculator[58] can be used. This algorithm is based on a biophysical model that can be used to predict the strength of an RBS sequence, allowing for the design of smaller degenerate libraries, typically of a dozen RBS variants that should cover a wide expression range[59,60]. As an alternative to RBS design based on biophysical models, high-throughput experimental RBS characterization data can be used to train models to predict RBS activity, as was recently demonstrated for *E. coli*[61,62]. Overall, the relatively strong predictive power for RBS strengths, as well as the requirement for only a few mutations within a short sequence window, makes them an excellent target for efficient, genome-wide optimization of expression strengths in bacteria. RBS engineering also allows for independently varying the expression levels of individual genes that are encoded together in one bacterial operon under the control of a single promoter.

By contrast, in eukaryotic cell factories such as yeast, genes are generally transcribed individually, and no RBS motif is present. Usually, the eukaryotic ribosome binds at the capped 5′ end of a transcript, then scans along until it encounters the start codon within the Kozak sequence to initiate translation[63]. Especially for yeast, predictive tools and standardized 5′UTR libraries are emerging based on secondary mRNA structures, the Kozak sequence motif and other internal regulatory elements controlling translation initiation. These advances can facilitate targeted system-wide tuning of expression in eukaryotes with 5′UTR libraries[64].

Another prominent target for modification is the protein coding sequence (CDS). The CDS can also be targeted for expression tuning by editing the codon usage of a gene (for example, codon optimization or codon harmonization). Several algorithms for codon optimizing are available[65], but editing the entire length of genes is often not practical, and also mostly not required for a system-wide approach. Targeting a shorter sequence like the promoter, the RBS or the influential start of the CDS can therefore be more attainable in system-wide genome engineering[66].

Apart from tuning expression levels, the CDS can be mutated to alter the amino-acid sequence of the encoded protein, for example to modulate enzyme properties such as catalytic rates or substrate or co-factor specificities. CDS sequences are often optimized via directed evolution strategies through mutagenesis of the (complete or partial) CDS, combined with screening or selection for desired enzyme properties. Directed evolution has, for example, enabled new-to-nature enzyme activities, such as the enzymatic formation of carbon–silicon bonds[67]. However, the variant space of all amino-acid substitutions (that is, 19 amino acids per position) in a single protein is astronomical. Hence, the system-level simultaneous engineering of proteins requires a strict selection of specific amino-acid residues, which are ideally mutated to only one or a few alternatives to find a compromise between minimizing the number of variants and covering potentially relevant residues. Available data on protein structures or previous characterizations or directed evolution studies on a specific protein can be used to limit the number of targets. Such an approach can also be used for the targeting of regulators. For example, targeted amino-acid modifications covering multiple regulator proteins have been performed in high-throughput metabolic engineering efforts to improve substrate or product tolerance. In one study, tolerance to hydrolysates from pretreated lignocellulosic biomass was engineered in *E. coli*[25]. The tolerance to acetate and furfural in the hydrolysate was improved by editing several key amino acids (mostly DNA-binding residues) of 28 genes, mostly regulators. This approach has already led to library sizes of >40,000 possible mutants, emphasizing the challenge of co-optimizing multiple CDS sequences in parallel. In another study, tolerance to the product styrene was improved through a similar approach, in which 54 regulators and transcription factors were edited, with >80,000 mutations tested in parallel[26]. Genome editing in both these studies was achieved through iterative CRISPR-Enabled Trackable genome Engineering (iCREATE), which will be described in more detail below.

Available and developing genome editing tools

Recent advances in DNA-editing techniques launched a new era in which precise genetic manipulation has become more efficient and feasible across entire genomes in an increasing number of microbial species. This transformation was primarily made possible by the development of recombination-mediated genetic engineering (recombineering)[6] and CRISPR-Cas-based genome editing[4]. In recent years, a range of genetic tools have been developed based on these two approaches, leading to several tools suitable for high-throughput genome editing for metabolic engineering (Fig. 2 and Table 1). Of general importance for system-wide metabolic engineering is the targeting of multiple targets in a population in parallel. In addition, it is important that multiple targeted edits can be combined within a single cell, which we refer to as multiplexing. Note that parallel targeting of multiple single edits in a population can also be referred to as multiplexing, but we will not use this term for that goal in this Review. Several modern genome editing tools allow for multiplexing by allowing for simultaneous targeting of two or more loci during one editing round in a single cell. Alternatively, multiplexing is achieved by performing iterative editing rounds (typically short rounds of a few days), by which multiple mutations can be accumulated in a single cell over time. A key parameter for both parallel and multiplex genome editing is the editing efficiency. This efficiency is usually defined as the percentage of edited cells harbouring a defined type of edit within the total population after one editing round.

Genome-wide editing based on recombineering tools

Recombineering is one of the most powerful tools that has been added to the systems metabolic engineering toolbox in recent decades, although currently it is available only for a few bacterial hosts. Recombineering harnesses the activity of single-stranded DNA annealing proteins (SSAPs), originating from bacteriophages, to anneal an exogenous DNA template to its partially complementary genomic or plasmid target and thus integrate modifications[68]. Recombineering with double-stranded (ds) DNA templates is a useful method for the insertion of large DNA fragments, up to hundreds of kilobases, which require only short homology flanks (~50 bases), into specific target sites in the genome. This method is routinely used in *E. coli* for making knockouts and insertions of genes, complete pathways and even entire chromosomal segments[69,70], and has substantially accelerated metabolic engineering projects in *E. coli* during the past two decades. However, dsDNA recombineering is generally not suitable for paralllel and multiplex genome editing due to the low frequency of edited cells (0.01–0.001%) in the population after recombineering[71].

In recent years, recombineering based on single-stranded DNA (ssDNA) templates has developed into a transformative tool for systems metabolic engineering, by enabling parallel and multiplex editing in multiple bacterial species[6,72,73]. ssDNA recombineering is typically performed with DNA oligos of 90 bases, with 1–30

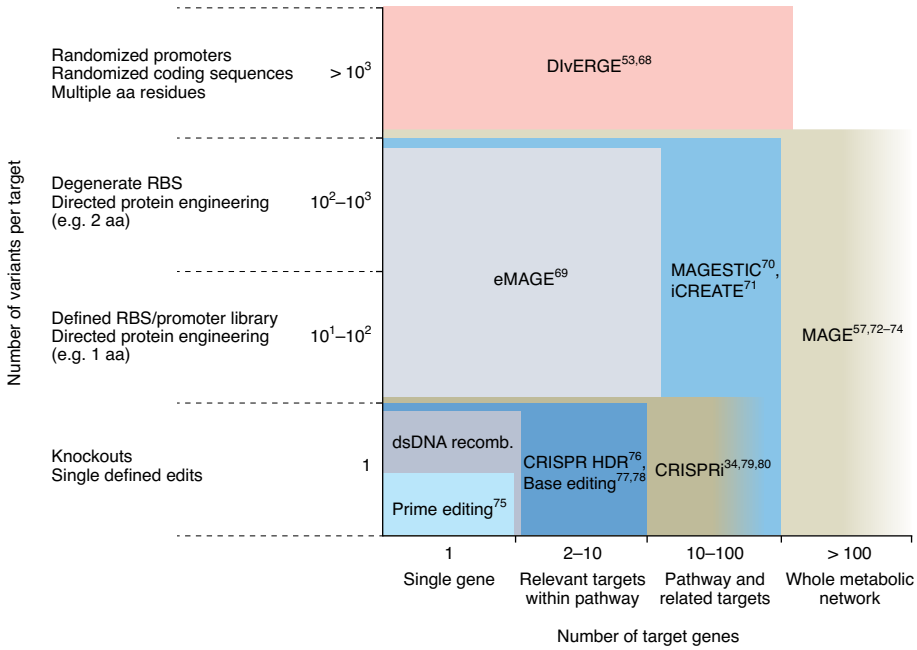

Fig. 2 | Editing ranges of different multiplex engineering tools. Different target types and number of target genes result in different ranges of possible mutant numbers, which can be achieved by different types of genome editing techniques. The *y* axis gives the number of genetic variants that approximately need to be generated per gene (or operon for prokaryotic promoters) for different target types. The *x* axis indicates the range of numbers of genes (or operons in the case of targeting prokaryotic promoters) that need to be targeted to perform metabolic engineering at different scales. The potential size of a strain library containing all possible combinations of mutations for a certain engineering scenario can be determined by the number of variants per target to the power of the number of target genes/operons. For example, by simultaneously targeting the RBSs of five genes with RBS libraries of ten RBS variants each, 10^5 combinations are possible. Coloured boxes indicate the typical coverage that can be provided in this space by specific genetic tools that are discussed in the text and in Table 1. The space covered by each tool is an approximation based on their currently demonstrated (maximum) potential in referenced metabolic engineering studies. Faded-gradient boxes indicate the expected potential of some tools based on demonstrated potential in non-metabolic engineering applications. Additionally, note that most tools can cover a specific mutational space in a population, so full mutational space is not feasible in a single cell. DIvERGE, directed evolution with random genomic mutations; MAGE, multiplex genome engineering; eMAGE, eukaryotic MAGE; dsDNA recomb., double-stranded DNA recombineering; CRISPR, clustered regularly interspaced palindromic repeats; CRISPR HDR, CRISPR (assisted) homology-directed repair; CRISPRi, CRISPR interference; aa, amino acid; RBS, ribosome binding site.

mismatching bases located in the middle of the oligo[6,72]. Bacterial cells in which SSAPs have been expressed (for example, by pORTMAGE plasmids[73]) are electroporated with these oligos to allow for incorporation of mutations during DNA replication (Fig. 3a). The electroporation with these oligos can be repeated in iterative cycles to accumulate mutations in individual cells. This type of iterative, ssDNA recombineering is commonly referred to as multiplex automated genome engineering (MAGE[6]). In practice, this technique does not necessarily require automation and can also be performed manually with standard molecular laboratory equipment.

One of the earliest demonstrated applications of MAGE was the simultaneous, rational targeting of 24 genes (native and non-native) involved in the biosynthesis of the molecule lycopene in *E. coli*. This allowed the generation of a few strains with fivefold-improved lycopene production within three days[57]. More recent demonstrations of ssDNA recombineering with larger libraries of oligos have allowed for the targeting of thousands of targets, for example, targeting promoters and RBSs of all protein-encoding genes in *E. coli*[74].

Recent screening efforts identified SSAPs that also allow ssDNA recombineering in other bacterial hosts besides *E. coli*. SSAP and recombineering protocols based on single plasmids carrying all required recombineering machinery have been established for several popular metabolic engineering hosts, such as *Pseudomonas putida*[75] and *Corynebacterium glutamicum*[76]. However, the best editing efficiencies in the latter two hosts so far are 1–5%. Efficiencies of >10%, enabling efficient parallel and multiplex engineering, have already been demonstrated for some other industrial hosts, such as the lactic-acid bacteria *Lactococcus lactis* and *Lactobacillus reuteri*[77,78], as well as the emerging bioproduction host *Citrobacter freundii*[79]. In *E. coli*, efficiencies of up to 50% have been reached using the best-performing SSAP (CspRecT[79]) for single-base mutations in one round. This ultrahigh-efficiency recombineering provides opportunities to integrate many targeted mutations after just a few rounds of recombineering[6]. Efficient ssDNA recombineering will probably be expanded to more, relevant bacterial hosts through further large screens of SSAPs combined with improved mechanistic understanding of SSAPs and their interactions with the host genetic machinery[77,78].

A downside of MAGE is that the editable region is typically below ~30 bases[51,57,79], which may limit the ability to discover beneficial genotypes for longer window targets, such as larger promoter regions or a complete or partial coding sequence. This can be resolved by tiling such a region with multiple MAGE oligos. However, when tackling such larger regions, the number of potential mutations can be rather high. Therefore, the ssDNA recombineering variant DIvERGE (directed evolution with random genomic mutations) can be employed (Fig. 3b)[53]. DIvERGE employs a set of soft-randomized oligos to introduce tunable levels of mutations covering a target region. Soft-randomized oligos are custom oligo mixes that contain a percentage of degeneracy for specific user-defined bases (for example, at a given native base A, 94% of

Table 1 | Summary of key genome engineering tools for targeted, system-wide metabolic engineering

	Name	Target type(s)	Template and other key features	Demonstrated multiplexability (−, +/−, +)[a]	Demonstrated hosts relevant for metabolic engineering	Refs.
Recombineering	MAGE	Insertions and substitutions <30 bp, deletions <30 bp.	90-nt-long ssDNA oligos, requires organism-specific SSAP.	+ (i) +/− (s)	*E. coli*, *P. putida*, *C. glutamicum*, *L. lactis*, *L. reuteri*, *C. freundii*, *S. enterica*, *K. pneumonia*, *C. crescentus*, *L. rhamnosus*, *M. smegmatis*	57,73,78,79, 134,143–145
	eMAGE	Insertions and substitutions <30 bp, deletions <100 bp. Targets within ~20 kb from a co-selectable marker.	90-nt-long ssDNA oligos, requires organism-specific SSAP.	+ (i) + (s)	*S. cerevisiae*	81
	DIvERGE	Mainly substitutions, covering multi-kilobasepair regions via tiling.	90-nt-long soft randomized oligos with partially degenerate bases, requires functional SSAP.	+/− (s) + (i)	*E. coli*, *S. enterica*, *C. freundii*, *K. pneumoniae*	53,80
CRISPR-Cas combined with recombination	Cas9/Cas12a + yeast HR	Gene knockouts, gene insertions.	Linear dsDNA of 40 bp–1 kb.	+/− (s)	For example, *S. cerevisiae*, *K. lactis*, *A. nidulans*, *Y. lypolytica*, *K. phaffi*	86,87
	MAGIC	Deletions of 28 bp, as well as downregulations (CRISPRi) and upregulations (CRISPRa).	CRISPR array oligos cloned into plasmid libraries, integrated with CRISPRi and CRISPRa.	+ (i) +/− (s)	*S. cerevisiae*	89
	ReScribe/ CRAM/ CRMAGE	Insertions, substitutions and deletions ~6–20 bp.	90-nt-long ssDNA oligos + CRISPR array cloned into plasmid.	+ (i) +/− (s)	*E. coli*, *P. putida*	90,91,93
	(i)CREATE	Insertions, substitutions and deletions <35 bp.	Synthetic oligo (max 200 nt) cloned into plasmid libraries (including gRNA + repair template). Allows for trackability of edits after one round. Can be iterated (iCREATE).	+ (i)	*E. coli*	25,26,96,97
	MAGESTIC	Substitutions of 3–6 bp (synonymous codons).	Cloned oligo (170 nt) into plasmid libraries (including gRNA and barcode).	–	*S. cerevisiae*	98
CRISPR-Cas editing	Base editing	Substitutions (C to T/G/A, or A to G) for 5–7 bp.	No template, deaminase fused to dead Cas9.	+ (s)	For example, *E. coli*, *P. putida*, *C. glutamicum*, *B. subitilis*, *C. beijerinckii*, *R. sphaeroides*, *Streptomyces* sp.	99,100,102, 104,105
	Prime editing	Insertions ≤33 bp, deletions ≤97 bp, substitutions ≤2 bp.	Template integrated in gRNA (pegRNA), reverse transcriptase fused to nickase Cas9.	+/− (s)	*E. coli*	101,106
Control of transcription/ translation levels	CRISPRi	Promoters and coding sequences for down-tuning/knockout, 1–14 genes.	gRNA(s) expressed from plasmid (for multiplex in arrays) together with a dead Cas variant.	+ (s)	For example, *E. coli*, *S. cerevisiae*, *C. glutamicum*, *Synechocystis* PCC6803, *B. subtilis*, *Y. lipolytica*, *C. ljungdahlii*	34,86,107, 108,146
	CRISPRa	Promoters for upregulation, 1–5 genes.	gRNA(s) expressed from plasmid (for multiplex in arrays) together with a dead Cas variant.	+ (s)	For example, *E. coli*, *K. oxytoca*, *Y. lipolytica*, *S. cerevisiae*, *B. subtilis*	86,113,114
	RNAi/RAGE	Transcribed regions for knockdown (UTRs, coding regions), 1–3 genes in one cell.	Library of mRNA template reverse-transcribed into DNA and inserted into plasmidlibrary.	+ (s)	*S. cerevisiae*	109
	sRNA	Transcribed regions for knockdown (UTRs, start coding regions), 1–3 genes in one cell.	Synthetic short sequences (20–120 bp) to act as antisense RNA expressed from plasmids.	+ (s)	For example, *E. coli*, *C. acetobutylicum*, *C. glutamicum*, *P. putida*, *B. subtilis*, *Synechocysistis*	110,111,147

[a]+ indicates the tool allows for multiplex targeting of many targets in single cell, +/− for a limited number of targets and − does not allow for multiplexing by respectively (i) = iterative, (s) = simultaneous targeting protocols.

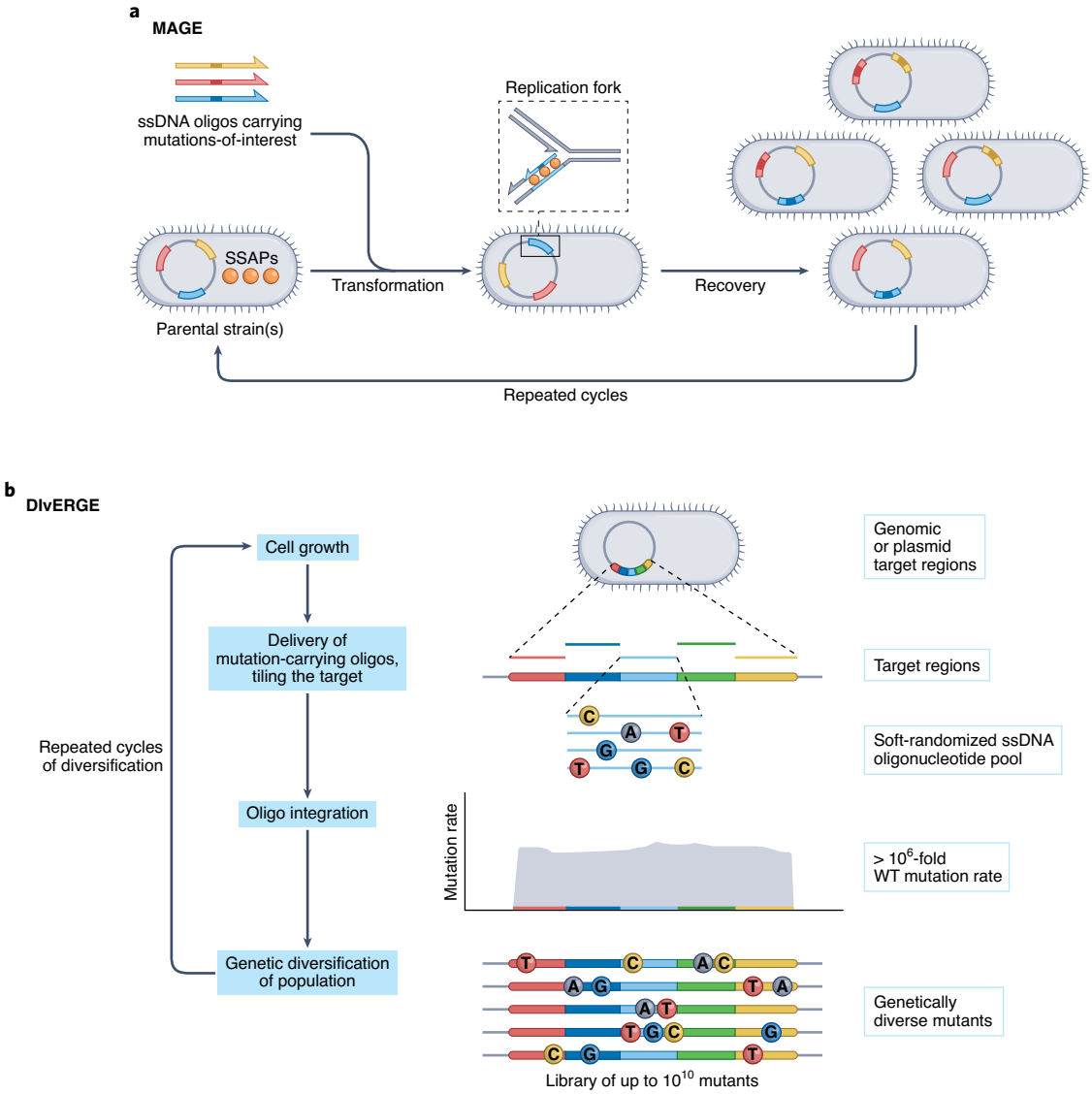

Fig. 3 | Single-stranded DNA recombineering allows for efficient multiplex genome engineering. a, MAGE allows for simultaneous editing of many targets in *E. coli* and some other bacteria by the introduction of mutagenic ssDNA oligonucleotides (left), which are incorporated at their genomic target sites during DNA replication (middle). Efficient integration of the oligonucleotides requires expression of suitable SSAPs. After a single round, a diversified population is generated (right), in which more mutations can be accumulated per cell by iterative rounds of MAGE. **b**, DIvERGE is a variant of ssDNA recombineering, which incorporates mutagenic soft-randomized oligos (that is, low percentage of degeneracy for each individual nucleotide-position) into the target region to increase the incidence of random mutagenesis at the site where the oligos anneal. Tiling of soft-randomized oligos can be used to cover a larger target site, such as coding sequences. Oligo integration results in a population with genetic diversity at the target site. Iterations can be used to combine multiple mutations in a single cell.

the oligo populations encodes wild-type A, 2% T, 2% C, 2% G). These oligos can tile the target's entire length, thereby maintaining the necessary homology for efficient integration and introducing a large range of distributed mutations (Fig. 3b). The DIvERGE approach has been utilized for studying mutations throughout antibiotic resistance genes, and for optimizing promoters and 5′UTRs for a synthetic (plasmid-based) regulatory circuit in *P. putida*[53,80]. DIvERGE could become a strong tool, for example to optimize the promoter regions or coding regions of enzymes. It is especially useful for cases for which no specific set of target mutations can be designed, and hence random mutagenesis of one or more specific regions in the genome is required.

Although the mechanism on which ssDNA recombineering is based only works in bacterial hosts, similar oligo-based multiplex genome engineering tools are under development for eukaryotes, such as *S. cerevisiae*[81,82]. Eukaryotic multiplex genome engineering (eMAGE) using oligos has been developed for this yeast species by temporally disabling DNA repair and slowing down replication[81]. To increase editing efficiencies, a selectable marker gene was introduced in a region near the targeted genes, and the design of oligos was optimized to achieve ultrahigh editing efficiencies of up to 90% (ref. [83]). However, a major limitation of eMAGE that is hampering genome-wide optimization is that efficient multiplex editing is only possible for a region covering ~20 kb in the proximity of the selectable marker,. Still, eMAGE allowed for the multiplex editing and optimization of a heterologous beta-carotene biosynthetic pathway in *S. cerevisiae*, for which all genes were co-localized in close proximity to a selectable marker[81].

CRISPR-Cas-aided genome editing combined with recombination

The programmable Cas nucleases derived from the natural, bacterial CRISPR-Cas defence systems have led to a revolution in the genome editing field in the past decade. A diverse range of CRISPR-Cas nuclease-based tools have been developed for genome engineering in both prokaryotic and eukaryotic cells, leading to many possible applications, including the optimization of next-generation cell factories[4,84,85]. DNA-targeting CRISPR-Cas nucleases such as Cas9 and Cas12a are very effective in many cell types to generate targeted, double-stranded breaks. After delivery (usually by expression from a plasmid), these nucleases are targeted to selected genome sequences by guide RNAs (gRNAs), which contain a variable sequence of ~20 nucleotides that can base-pair with a complementary target sequence. However, target sites need to be adjacent to a protospacer adjacent motif (PAM) sequence, such as the typical 5′-NGG PAM-motif for Cas9 or 5′-TTTV for Cas12a.

After the Cas nuclease generates a double-strand break in the DNA, in some organisms the break can be repaired by non-homologous end joining (NHEJ). NHEJ typically leads to random insertions or deletions (indels), which can, for instance, knock out genes by creating frame shifts, or to disrupt regulatory sequences.

However, to generate specific edits, the use of Cas nucleases should be combined with the homology-directed repair (HDR) system (Fig. 4a). In such cases, an exogenous ssDNA or dsDNA repair template should be supplied together with the Cas/gRNA complex. The repair fragment contains the desired edits between flanking regions that are homologous to the genomic target site, allowing for integration through homologous recombination by a HDR system. This recombination modifies the genomic target sequence, thus preventing subsequent recognition and cleavage by the Cas nuclease, and leading to the survival of cells with edited genomes.

Cas-based multiplexing requires highly efficient homologous recombination, which is not present naturally in many industrial hosts, including most bacteria. However, the most widely used industrial yeast species, *S. cerevisiae*, natively harbours highly efficient homologous recombination activity that can well support high-throughput genome editing. In *S. cerevisiae*, several studies have achieved five to ten edits in one editing round[86,87,88]. In addition to making multiple edits simultaneously in a single cell, the efficient homologous recombination combined with Cas9 has allowed for parallel modification of many targets in *S. cerevisiae*. Such approaches have been demonstrated in *S. cerevisiae* to generate diverse large libraries with 10^4 to 10^5 different targets, and by iteration of such workflows many mutations can be accumulated in a single strain. An example of this was the iterative, CRISPR-mediated modification of $>10^5$ targets in *S. cerevisiae* (MAGIC approach) to increase its tolerance to furfural, an inhibitor found in pretreated lignocellulosic substrates[89].

However, in most other organisms, including typical bacterial cell factories, the native homologous recombination machinery is not efficient enough to allow for efficient repair during simultaneous, multiplex CRISPR-Cas genome editing. Still, by combining CRISPR-Cas counter-selection with recombineering in bacteria, the apparent editing efficiency can be increased up to 100%. Cas9 nucleases and ssDNA recombineering (MAGE) were integrated to more efficiently target multiple shorter regions in *E. coli* in several CRISPR-assisted MAGE approaches, including CRAM[90] and CRMAGE[91]. Using CRMAGE, editing efficiencies of up to 98% have been achieved in a single round of recombineering-and-CRISPR-selection. By applying only a few cycles of CRISPR-assisted MAGE, the production of riboflavin in *E. coli* was enhanced via multiplex RBS engineering[92]. Integration of CRISPR and recombineering in more species is expected to improve high-throughput genome editing in several other promising cell factories in the near future, as recently demonstrated for *P. putida*[93].

To introduce larger mutations (≫30 bp) dsDNA recombineering is generally used, but this results in very low frequencies of edited cells. This necessitates the use of either a selectable (antibiotic) marker or extensive screening of many cells, which prohibits fast and efficient high-throughput editing. However, by combining dsDNA recombineering with CRISPR-Cas in *E. coli*, multiple large edits (up to three), without the need for selectable markers, can be made relatively efficiently[94,95]. This can speed up rational metabolic engineering of larger targets in *E. coli*, and possibly also in other hosts that are amenable for recombineering.

The trackability of genomic edits is another challenge during multiplex genome editing of many targets in parallel. When many sites are targeted throughout the genome, tracing of the edits generally requires relatively costly whole-genome sequencing or laborious screening polymerase chain reactions (PCRs). This has been tackled in the CRISPR-assisted recombineering platform CREATE (CRISPR-Enabled Trackable genome Engineering; Fig. 4e)[96]. CREATE is based on libraries of plasmids, encoding both the gRNA and the (short) dsDNA repair template, which can be synthesized as libraries of oligos, which are cloned into the plasmid. After one round of editing, the plasmids that created the mutation are still in the mutant strains, and their unique gRNA + repair template sequences can serve as a barcode to easily identify the target mutation in well-performing cells after screening or selection. The CREATE platform, and the use of it in iterative cycles (iCREATE), has been demonstrated to be very powerful for the parallel, genome-wide editing of $>10^5$ mutations in multiple metabolic engineering projects in *E. coli*, including, for example, the earlier mentioned tolerance to furfural in lignocellulosic substrate and the toxic product styrene[25,26,97]. A similar trackable, CRISPR-based editing strategy, called MAGESTIC, was also recently developed for *S. cerevisiae*[98].

Genome editing by CRISPR-Cas without recombination

A general challenge for all Cas-nuclease methods based on the generation of dsDNA breaks is the need for their efficient repair through efficient homologous recombination or recombineering. Such efficient repair is often not available or easy to implement in many metabolic engineering hosts. Recent developments in CRISPR-based base editing[99,100] and prime editing[101] circumvent the limitations of dsDNA break formation, while allowing for inheritable genomic edits. Base editing relies on fusing a nuclease-deficient non-cleaving 'dead' Cas9 (dCas9) to an accessory enzyme that chemically alters DNA to introduce the desired change (Fig. 4c). For example, fusing the cytidine deaminase PmCDA1 to dCas9 leads to efficient C-to-T mutagenesis at the target site[102]. In addition to C-to-T cytosine base editors (CBEs), A-to-G adenine base editors (ABEs) have been developed, thus enabling several nucleotide substitutions. However, this tool only allows substitution in short windows—and no deletions or insertions[88,103].

Recently, CBE base editing was used to diversify RBS sequences in the two metabolic engineering workhorses *Corynebacterium glutamicum* and *Bacillus subtilis* using the BETTER (Base Editor-Targeted and Template-free Expression Regulation) protocol[104]. In this approach, a synthetic starter RBS (containing a stretch of eight Gs) preceded by a PAM-sequence was integrated in front of each target gene. This preparative, multiple integration step was not efficiently multiplexed, making this approach rather laborious. Nonetheless, the CBE editor could efficiently target all these synthetic RBSs simultaneously, diversifying them by converting Gs to As, leading to diverse translation initiation strengths. Another recently developed base editing protocol for *P. putida* allowed for multiplexing of knockouts and modifications for system metabolic engineering[105].

A very recently developed CRISPR-tool is prime editing, which allows for more versatile, precision genome editing for all types of

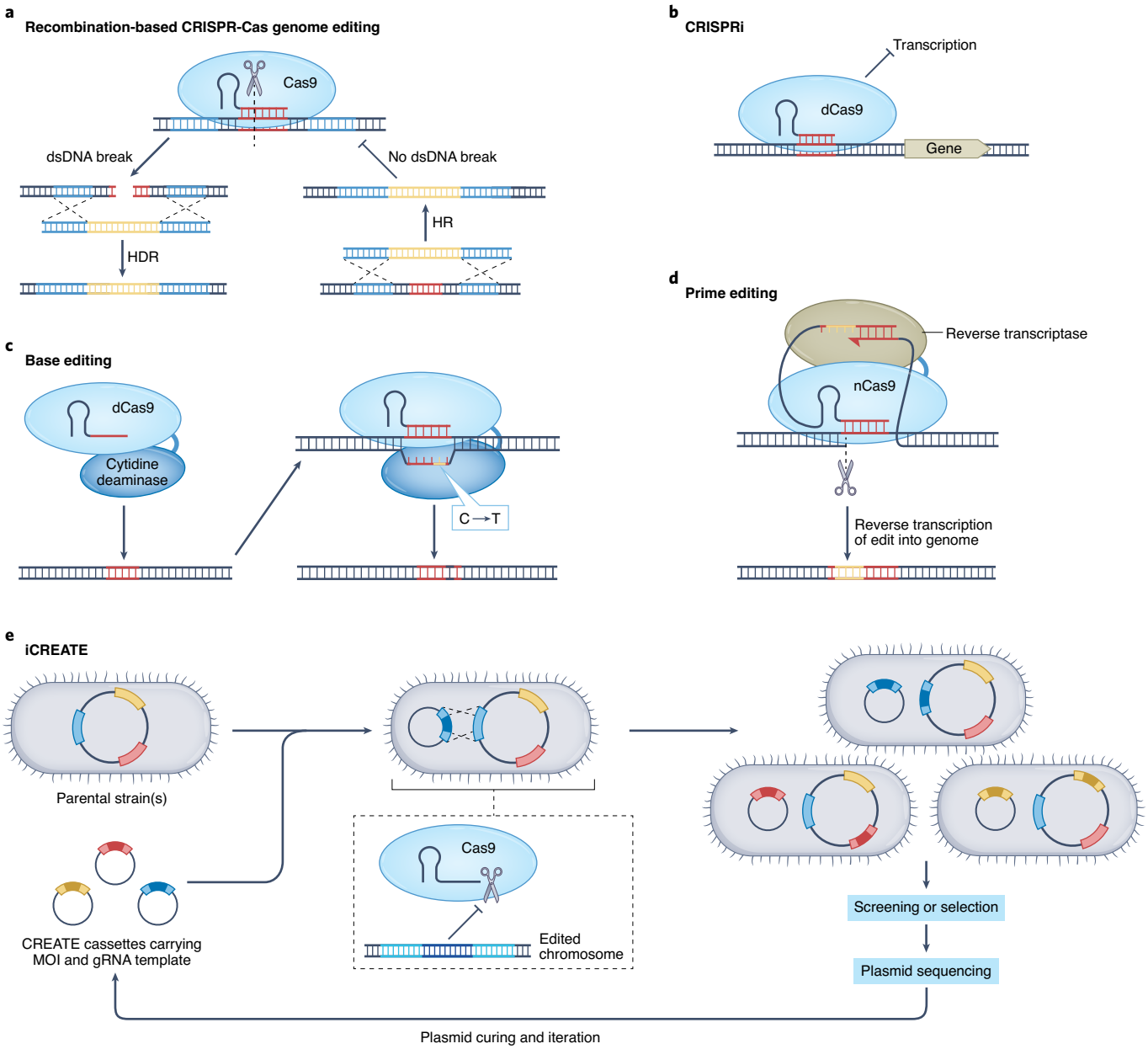

Fig. 4 | CRISPR-Cas based genetic tools. a, CRISPR-Cas genome editing based on double-stranded break generation by a gRNA-guided Cas nuclease. For example the CRISPR-Cas nuclease Cas9 targeting can facilitate genome editing in two ways. First, induction of a dsDNA break can trigger HDR with a repair template, leading to insertion or deletion of a custom sequence. Alternatively, homologous recombination (HR) with the template can occur spontaneously, before Cas-induced double-stranded breaks happen, and hence allow for the counter-selection of non-edited cells in a population. **b**, CRISPRi is based on a non-cleaving dead Cas (dCas) variant, which binds upstream or within the gene of interest (GOI) and blocks transcription. **c**, Base editors consist of dCas fused to a DNA-editing protein like cytidine deaminase. The gRNA guides dCas and cytidine deaminase to a selected target site where the latter induces SNPs in the displaced strand, after which a mismatch repair system can also adjust the other strand. **d**, Prime editing uses a nickase Cas (nCas) variant fused to a reverse transcriptase (RT) to first induce a single-stranded nick in the displaced strand. A custom sequence is then reverse-transcribed into the cleavage site, using the prime editing guide RNA (pegRNA) as a repair template. **e**, iCREATE uses massively parallel DNA synthesis to construct a large library of plasmid-encoded CREATE cassettes carrying a gRNA and repair template with mutation of interest (MOI) for a large variety of targets. This cassette library is introduced in a parental strain expressing Cas9 to construct a large strain library in which individual cells carry a single plasmid from the library, which results in editing of a single target or cell death by CRISPR-Cas counter-selection. The resulting diversified strain library can be screened or selected for the desired phenotype, and enrichment of MOIs can be determined by sequencing the CREATE cassettes.

nucleotide substitutions as well as short deletions and insertions, encompassing larger editing windows of >30 bp (ref. [103]). Prime editors were generated by fusing an engineered Cas9 nickase (nCas9), capable of introducing a single-strand break in dsDNA, to a reverse transcriptase (Fig. 4d). Prime editing works through reverse transcription (RNA to ssDNA) of a desired edit into the genome at the nicked target site, using a modified prime editing RNA (pegRNA) as repair template. The pegRNA contains both the guide sequence complementary to the target site, the reverse transcriptase template carrying the mutation of interest, as well as a

short primer binding site where reverse transcription starts. Prime editing has already been performed successfully in a few eukaryotic and prokaryotic organisms[103,106]. Recently, prime editing has been performed in *E. coli*, resulting in single-base deletions at 40% efficiency. This study also described simultaneous editing of two single base targets, but the multiplex editing efficiency was below 1%. With further development to improve editing efficiencies, this emerging tool may hold promise for genome-wide metabolic engineering in a broad range of organisms[106].

Tools for multiplexed gene expression control without editing

There are also a suite of multiplex tools that do not lead to edits at the genome level, but rather modify expression at the transcription or translation level. This includes the relatively well-established CRISPR interference (CRISPRi) tools (Fig. 4b). CRISPRi harnesses catalytically impaired dead Cas variants (for example, dCas9 or dCas12a), which remain capable of sequence-specific DNA-binding. They can be targeted to several specific promoter or coding regions on the DNA in a cell in parallel to repress transcription by blocking the RNA polymerase complex. This technique has already found some applications in multiplex, rational metabolic engineering of up to ten targets simultaneously[86,107], and a recent metabolic engineering study in *P. putida* successfully targeted 14 rationally selected genes simultaneously for knockdown[34]. CRISPRi seems a potentially powerful technique, but the design and control of guide expression to finely tune transcription rates remains challenging[108]. The CRISPRi technique does not lead to inheritable genome edits, which can be beneficial as it allows for temporary control, for example, in either the growth or production phase. However, as a potential disadvantage, it can also be escaped relatively easily, for example, by a single mutation inactivating the dCas variant.

Other knockdown strategies, such as RNA interference (RNAi), have some of the same benefits and drawbacks as CRISPRi. RNAi is based on the silencing of mRNA translation by the binding of a complementary RNA. RNAi-assisted genome evolution (RAGE) has been used to identify targets for engineering improved acetic-acid tolerance in *S. cerevisiae*. RAGE employs an RNA library constructed from a fragmented host genome for iterative rounds of RNAi, combined with high-throughput screening or selection[109]. However, because targets are randomly generated, RAGE is more suitable for identifying targets than for rational engineering. Alternatively, rationally designed synthetic small regulatory RNAs (sRNAs) have, for instance, been used for targeted knockdown of genes involved in the tyrosine pathway in *E. coli* to improve tyrosine titres[110]. sRNAs have been demonstrated to work in multiple bacterial species, but so far the number of targets tackled simultaneously in one cell has mostly been limited to two[111].

As an alternative to gene knockdown, a less developed, but promising technique based on dCas to activate gene expression is CRISPR activation (CRISPRa). In this case, dCas is fused to transcription activation domains to recruit them to specific promoters in the genome and stimulate recruitment of the RNA polymerase, hence activating transcription of the downstream gene(s)[86]. This has been successfully developed in several hosts, including *E. coli* and recently *P. putida*[112,113].

Screening and selection strategies

As more powerful high-throughput genome editing tools have become available, larger diversified strain libraries can be constructed, and equally powerful high-throughput screening or selection strategies are needed to identify desired phenotypes.

Traditional screening methods (Fig. 5a) rely on the isolation and culturing of individual strains and quantification of a product molecule by analytical techniques such as spectroscopy, high-pressure liquid chromatography (HPLC) and gas chromatography–mass spectrometry (GC-MS)[114]. Technological advances have vastly increased the throughput of these methods via automation and massive parallelization of sample processing, and miniaturization of the culture scale[115,116].

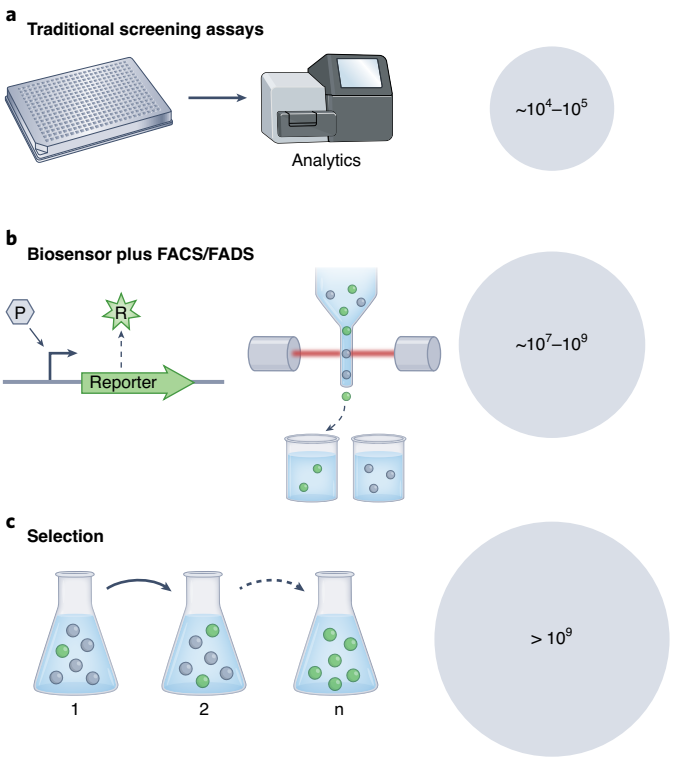

Fig. 5 | Scales of different screening and selection strategies. a, Traditional screening assays involve individual strain isolation, cultivation and analysis (for example, spectroscopy, HPLC and (GC-)MS) and are limited to throughputs of ~10^4–10^5 variants per day when using a commercially available automated colony-picking machine and high-throughput analytics equipment. **b**, Reporter-coupled screening through fluorescence-activated single cell or droplet sorting does not require isolated strains cultivation and can reach up to 10^7–10^9 cells per day using a commercially available cell-sorting device. P, product molecule; R, fluorescent reporter protein. **c**, Growth or reporter-coupled selection requires continuous cultivation over a longer period and is highly dependent on the growth rate advantage of a selectable trait. Therefore, it is difficult to quantify as a daily throughput, but can ultimately be used for massive populations (>10^9).

Spectrophotometric assays are commonly used as high-throughput screening approaches for measuring product titres in microtitre plates. Automated colony-picking and sample-processing systems have enabled high-throughput microtitre plate assays of up to ~3,000 colonies per hour[117]. Some products can be detected with ultraviolet–visible (UV–vis) spectrometry (such as β-carotene, lycopene and apigenin), but unfortunately this technique suffers from poor specificity and limited sensitivity[115]. Some products have intrinsic fluorescence (for example, riboflavin, thiamine, pyridoxine) that can be detected with high sensitivity and specificity by fluorescence spectrometry[118]. However, most products of interest cannot be detected directly by colorimetric or fluorescent spectrometry, as most products of interest are not sufficiently spectroscopically active[119]. Mass spectrometry (MS) does not have this requirement, and is a versatile and widely applied screening method for quantifying product titres because of its high selectivity, sensitivity and ability to detect a wide range of compounds[120]. Ultrahigh-throughput MS platforms have been established that reduce preparation times and enable the measurement of multiple

samples per second. These include acoustic ejection mass spectrometry (AEMS)[121] and acoustic mist ionization MS (AMI-MS)[122]. However, despite these technological advances, high-throughput MS remains limited by the throughput of microtitre plate-based culturing and sample processing.

Single-cell-sorting methods have eliminated the need for individual strain cultivation by directly analysing single cells from a diversified population. Active-cell-sorting technologies use principles of flow cytometry to separate, detect and analyse individual cells and subsequently sort them based on various characteristics (for example, fluorescence, size and shape) using electric, magnetic, acoustic or optical forces to displace the cells[123]. Modern fluorescence-activated cell sorting (FACS) techniques allow for an extremely high-throughput of up to 50,000 cells per second[123]. A limitation of FACS is that it cannot be used to select high-producing cells when products are excreted, as is often the case in cell factories. Droplet-based microfluidics sorting provides a solution to this problem by compartmentalizing the individual cells in aqueous droplets emulsified in oil[119,124], but sorting rates are substantially lower than for FACS, with ~200 droplets per second[125]. For example, fluorescence-activated droplet sorting (FADS) was recently applied to enhance pyruvic-acid production by *Candida glabrata*[126]. Because pyruvic acid is not inherently fluorescent, a pH-sensitive fluorescent protein (pHluorin) was used to link pyruvic-acid levels secreted by the cells to a fluorescent signal. This illustrates an important challenge of flow cytometry and FACS/FADS: most products cannot be directly detected by fluorescent signals or cell morphological properties.

To circumvent issues with products that are not detectable by high-throughput analytics like FACS, biosensors can be used to couple the presence of the product to the expression of an easily detectable reporter (Fig. 5b). There are many mechanisms for product–reporter coupling, including transcription factor-based biosensors, nucleic acid-based biosensors and fluorescent detector proteins that gain fluorescence upon binding of the product[116,127–129]. Diverse biosensors have been used successfully for the detection of a variety of metabolites in *E. coli*, such as deoxyviolacein[130], malonyl-CoA[131] and mevalonate[132] biosensors. However, suitable biosensors are not available for every desired product, and the construction of specific biosensors has proven highly challenging. For further reading on other biosensors we recommend a comprehensive review[114].

An advantageous versatility of biosensors is that, besides linking product formation to the expression of a fluorescent or colorimetric reporter gene, they can also be used in combination with a selectable marker such as antibiotic resistance[133]. Selection of strains based on the expression of a selectable marker can form a higher-throughput alternative to screening, while also removing the need for expensive detection and sorting devices. A challenge of biosensor-coupled selection is the high selective pressure for false-positive 'cheater mutants', as mutations can occur that decouple expression of the selectable marker gene from the product formation[133]. Successful use of a selectable marker coupled to biosensors has been demonstrated for the high-throughput optimization of naringenin and glucaric acid production in *E. coli* (using MAGE). To cope with the escaper issue, this work coupled the biosensor to both a negative and a positive selection marker, which were alternately selected for in iterations, eliminating the escapers[134].

In specific cases, metabolic engineering goals can be directly aligned with an improved growth phenotype without the assistance of a biosensor (Fig. 5c). For example, selection can provide a higher-throughput alternative to screening when high tolerance to a toxic substrate, product or intermediate metabolite is required, because improved tolerance will confer improved fitness in the presence of the toxin. In addition, in cases where the engineered pathway leads to production of an essential metabolite, auxotrophy can be artificially created by knockout of alternative pathways, creating a growth-coupling between the engineered pathway and the production of these metabolites. This approach is especially useful for engineering substrate assimilation pathways[135]. Several engineering efforts for introducing heterologous substrate utilization have used such growth-coupled approaches, such as the engineering of one-carbon assimilation pathways. However, so far, these studies have most relied on low-throughput rational engineering of a few targets within the pathway, combined with ALE to optimize host metabolism[38,136,137]. Still, some studies have already shown the power of growth selection for relatively small sets of MAGE-generated, targeted mutations in substrate utilization pathways[136,138].

Another promising application of engineered auxotrophy is in Syntrophic Co-culture Amplification of a Production phenotype (SnoCAP)[139]. This elegant screening strategy is based on co-culture of the producer strain with a sensor strain, where both are co-dependent on an excreted metabolite from the other strain. The producer strain excretes the product of interest 'A', but is auxotrophic for essential metabolite 'B', and the sensor strain has the inverse phenotype (that is, auxotrophy for product 'A' but excretes metabolite 'B'). By skewing the required ratio of producer to sensor strain towards the latter, this approach allows screening for higher product titres than an individual cell requires to fullfill the flux demand for its own auxotrophy.

Conclusions and outlook

To realize the true promise of sustainable production by microorganisms, large improvements in both native and engineered metabolic networks are required at a system-wide scale. This will require (semi-) rational selection of a large number of diverse gene targets for optimization. In well-known metabolic engineering hosts, such as *E. coli* and *S. cerevisiae*, these targets may be identified based on an impressive amount of previously gained knowledge on their metabolic networks and their regulation. However, for other less-studied and emerging hosts for industrial production, acquiring such detailed knowledge will be time-consuming and costly. Still, genomics and other omics techniques, as well as metabolic modelling and machine learning, can probably provide suitable, semi-rational sets of potentially relevant targets without excessive effort. We believe that the semi-rational selection of targets in many cases will be more effective for strain improvement than only using the random ALE approach, as is now often used for genome-wide strain optimization.

Yet, to tackle a larger set of semi-rationally selected targets, efficient high-throughput genome editing tools are crucial. Fortunately, the current development of recombineering and CRISPR-Cas genome editing techniques allows for genetic engineering of a growing number of model and non-model (micro)organisms. The rapid emergence of CRISPR-tools in a broad range of bacterial and eukaryotic hosts is encouraging. However, classic CRISPR-Cas tools based on dsDNA cleavage are often not efficient enough for generating multiple edits simultaneously or even in iterative cycles. So far, in bacteria, highly efficient multiplex genome editing based on CRISPR-Cas nucleases is only well-established in *E. coli*, in which CRISPR-Cas is combined with recombineering. In some industrially relevant yeast species such as *S. cerevisiae*, CRISPR-Cas editing seems efficient enough for high-throughput iterative genome editing, as in this species native homologous recombination can be harnessed to efficiently incorporate DNA repair templates. However, for many other bacterial and eukaryotic hosts, CRISPR-Cas editing systems that make double-stranded breaks are not effective for high-throughput editing without establishing complementary, efficient recombineering systems. Encouragingly, recent screening expeditions to identify efficient phage SSAPs for recombineering in several bacteria have been successful[6]. Such efforts can probably identify SSAPs for efficient recombineering above 10% editing efficiency, to allow efficient genome editing via recombineering in more hosts of interest. In addition, further development of

CRISPR-Cas editing systems will allow tighter control of DNA cleavage activity and allow for more effective application as a powerful counter-selection tool in more hosts[140].

Alternatively, there are promising, emerging CRISPR-Cas tools that do not rely on dsDNA breaks and recombineering, but rather employ dead or nickase variants of the CRISPR-Cas nucleases. This includes CRISPR interference, base editing and emerging prime editing techniques, which do not require external DNA templates and should be explored more for multiplex editing in diverse hosts relevant for metabolic engineering.

Within the next decade, we expect that metabolic engineers will gradually gain greater freedom to choose a desired host for production, and rapidly apply the available genetic tools for efficient, genome-wide engineering. In addition, high-throughput strain engineering will benefit from the automation of iterative, genome engineering techniques, for example in emerging biofoundries in industry and academia[141,142].

Another factor that should be carefully considered when performing genome-wide strain optimization is the availability of high-throughput screening or selection strategies to identify desired phenotypes. Matching the number of mutants to be generated with the number that can be screened or selected is important to consider in the design phase of a system-wide metabolic engineering project.

Successful integration of the (semi-)rational selection of targets, high-throughput genome editing, and screening or selection approaches discussed in this Review provides a strong foundation towards directed, system-wide metabolic engineering. These developments seem crucial to move the boundaries of metabolic engineering towards economically feasible performance indicators, as well as to implement non-natural biosynthesis for products that could not be made by biology before. Overall, we anticipate that the developments discussed here could contribute to a revolution in the sustainable production of desired products by next-generation cell factories.

Received: 21 July 2021; Accepted: 28 July 2022;
Published online: 20 September 2022

References

1. Bailey, J. E. Toward a science of metabolic engineering. *Science* **252**, 1668–1675 (1991).
2. Casini, A., Storch, M., Baldwin, G. S. & Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.* **16**, 568–576 (2015).
3. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
4. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651 (2018).
5. Mougiakos, I., Bosma, E. F., Ganguly, J., van der Oost, J. & van Kranenburg, R. Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. *Curr. Opin. Biotechnol.* **50**, 146–157 (2018).
6. Wannier, T. et al. Recombineering and MAGE. *Nat. Rev. Methods Prim.* **1**, 7 (2021).
This review provides an overview on the mechanism, applications and optimal experimental strategies of recombineering and specifically MAGE, in diverse organisms.
7. Liao, J. C., Mi, L., Pontrelli, S. & Luo, S. Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat. Rev. Microbiol.* **14**, 288–304 (2016).
8. Claassens, N. J. et al. Harnessing the power of microbial autotrophy. *Nat. Rev. Microbiol.* **14**, 692–706 (2016).
9. Lee, S. Y. et al. A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2**, 18–33 (2019).
This review article gives a comprehensive overview of industrial production for many bio-based products including their metabolic routes, current performances and challenges.
10. Choi, K. R. et al. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol.* **37**, 817–837 (2019).
11. de Jong, E., Stichnothe, H., Bell, G. & Jørgensen, H. *Bio-Based Chemicals: a 2020 Update* (IEA Bioenergy, 2020).
12. Oxford Economics. *The Global Chemical Industry*: *Catalyzing Growth and Addressing Our World's Sustainability Challenges* (Oxford Economics, 2019); https://www.oxfordeconomics.com/recent-releases/the-global-chemical-industry-catalyzing-growth-and-addressing-our-world-sustainability-challenges
13. Blanco, J., Iglesias, J., Morales, G., Melero, J. A. & Moreno, J. Comparative life cycle assessment of glucose production from maize starch and woody biomass residues as a feedstock. *Appl. Sci.* **10**, 2946 (2020).
14. Naz, T. et al. Transformation of lignocellulosic biomass into sustainable biofuels: major challenges and bioprocessing technologies. *Am. J. Biochem. Biotechnol.* **16**, 308–327 (2020).
15. Fackler, N. et al. Stepping on the gas to a circular economy: accelerating development of carbon-negative chemical production from gas fermentation. *Annu. Rev. Chem. Biomol. Eng.* **12**, 439–470 (2021).
16. Cotton, C. A. R., Claassens, N. J., Benito-Vaquerizo, S. & Bar-even, A. Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* **62**, 168–180 (2020).
17. Wehrs, M. et al. Engineering robust production microbes for large-scale cultivation. *Trends Microbiol.* **27**, 524–537 (2019).
18. Lee, J. W. et al. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* **8**, 536–546 (2012).
19. Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* **33**, 1061–1072 (2015).
20. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56**, 1–16 (2019).
21. Wu, Y., Jameel, A., Xing, X. H. & Zhang, C. Advanced strategies and tools to facilitate and streamline microbial adaptive laboratory evolution. *Trends Biotechnol.* https://doi.org/10.1016/j.tibtech.2021.04.002 (2021).
22. Biz, A. et al. Systems biology based metabolic engineering for non-natural chemicals. *Biotechnol. Adv.* **37**, 107379 (2019).
23. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab. Eng.* **9**, 258–267 (2007).
24. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568 (2006).
25. Liu, R. et al. Directed combinatorial mutagenesis of *Escherichia coli* for complex phenotype engineering. *Metab. Eng.* **47**, 10–20 (2018).
A demonstration of the use of iCREATE to target ~40,000 mutations across 30 diverse genes in parallel, to optimized C5/C6 sugar co-utilization and tolerance to furfural and acetate in *E. coli*, which occur in lignocellulose hydrolyates.
26. Liang, L. et al. Genome engineering of *E. coli* for improved styrene production. *Metab. Eng.* **57**, 74–84 (2020).
27. Watson, E., Yilmaz, L. S. & Walhout, A. J. M. Understanding metabolic regulation at a systems level: metabolite sensing, mathematical predictions and model organisms. *Annu. Rev. Genet.* **49**, 553–575 (2015).
28. Banos, D. T., Trébulle, P. & Elati, M. Integrating transcriptional activity in genome-scale models of metabolism. *BMC Syst. Biol.* **11**, 81–90 (2017).
29. Nielsen, J. Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).
30. Stalidzans, E., Seiman, A., Peebo, K., Komasilovs, V. & Pentjuss, A. Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochem. Soc. Trans.* **46**, 261–267 (2018).
31. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987 (2015).
32. Burgard, A. P., Pharkya, P. & Maranas, C. D. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).
33. Von Kamp, A. & Klamt, S. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat. Commun.* **8**, 15956 (2017).
34. Banerjee, D. et al. Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. *Nat. Commun.* **11**, 5385 (2020).
35. Sánchez, B. J. et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* **13**, 935 (2017).
36. Kim, O. D., Rocha, M. & Maia, P. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. *Front. Microbiol.* **9**, 1690 (2018).
37. Lee, Y., Lafontaine Rivera, J. G. & Liao, J. C. Ensemble modeling for robustness analysis in engineering non-native metabolic pathways. *Metab. Eng.* **25**, 63–71 (2014).

38. Chen, F. Y. H., Jung, H. W., Tsuei, C. Y. & Liao, J. C. Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol. *Cell* **182**, 933–946 (2020).
This article reports the kinetic-modelling informed design, rational construction, ALE and realization of full growth on methanol in *E. coli* via the RuMP pathway.
39. Zhou, H., Vonk, B., Roubos, J. A., Bovenberg, R. A. L. & Voigt, C. A. Algorithmic co-optimization of genetic constructs and growth conditions: application to 6-ACA, a potential nylon-6 precursor. *Nucleic Acids Res.* **43**, 10560–10570 (2015).
40. Xu, P., Rizzoni, E. A., Sul, S. Y. & Stephanopoulos, G. Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth. Biol.* **6**, 148–158 (2017).
41. Young, E. M. et al. Iterative algorithm-guided design of massive strain libraries, applied to itaconic acid production in yeast. *Metab. Eng.* **48**, 33–43 (2018).
42. Jervis, A. J. et al. Machine learning of designed translational control allows predictive pathway optimization in *Escherichia coli*. *ACS Synth. Biol.* **8**, 127–136 (2019).
43. Zhou, Y. et al. MiYA, an efficient machine-learning workflow in conjunction with the YeastFab assembly strategy for combinatorial optimization of heterologous metabolic pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* **47**, 294–302 (2018).
44. Zhang, J. et al. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. *Nat. Commun.* **11**, 4880 (2020).
Good illustration of integrated system metabolic engineering approaches in yeast, based on modeling to identify five gene targets, Cas9-based editing of promoters variants for these genes, recording data with a tryptophan biosensor and using generated data for machine learning for further improve tryptophan production.
45. Lawson, C. E. et al. Machine learning for metabolic engineering: a review. *Metab. Eng.* **63**, 34–60 (2021).
46. Volk, M. J. et al. Biosystems design by machine learning. *ACS Synth. Biol.* **9**, 1514–1533 (2020).
47. Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Machine learning applications in systems metabolic engineering. *Curr. Opin. Biotechnol.* **64**, 1–9 (2020).
48. Li, F. et al. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* **5**, 662–672 (2022).
49. Urtecho, G. et al. Genome-wide functional characterization of *Escherichia coli* promoters and regulatory elements responsible for their function. Preprint at https://www.biorxiv.org/content/10.1101/2020.01.04.894907v1 (2020).
50. La Fleur, T., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. Preprint at https://www.biorxiv.org/content/10.1101/2021.09.01.458561v1 (2021).
51. Wang, H. H. et al. Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* **9**, 591–593 (2012).
52. Yona, A. H., Alm, E. J. & Gore, J. Random sequences rapidly evolve into de novo promoters. *Nat. Commun.* **9**, 1530 (2018).
53. Nyerges, Á. et al. Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proc. Natl Acad. Sci. USA* **115**, E5726–E5735 (2018).
54. Cetnar, D. P. & Salis, H. M. Systematic quantification of sequence and structural determinants controlling mRNA stability in bacterial operons. *ACS Synth. Biol.* **10**, 318–332 (2021).
55. Nieuwkoop, T., Finger-Bou, M., van der Oost, J. & Claassens, N. J. The ongoing quest to crack the genetic code for protein production. *Mol. Cell* **80**, 193–209 (2020).
56. Zelcbuch, L. et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **41**, e98 (2013).
57. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
This landmark study established the principles of MAGE in *E. coli* and demonstrated its strong potential for targeted engineering of 24 genes involved in lycopene biosynthesis.
58. Reis, A. C. & Salis, H. M. An automated model test system for systematic development and improvement of gene expression models. *ACS Synth. Biol.* **9**, 3145–3156 (2020).
59. Farasat, I. et al. Efficient search, mapping and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10**, 731 (2014).
60. Jeschek, M., Gerngross, D. & Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.* **7**, 11163 (2016).
61. Bonde, M. T. et al. Predictable tuning of protein expression in bacteria. *Nat. Methods* **13**, 233–236 (2016).
62. Höllerer, S. et al. Large-scale DNA-based phenotypic recording and deep learning enable highly accurate sequence-function mapping. *Nat. Commun.* **11**, 3551 (2020).
63. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).
64. De Nijs, Y., De Maeseneire, S. L. & Soetaert, W. K. 5′ untranslated regions: the next regulatory sequence in yeast synthetic biology. *Biol. Rev.* **95**, 517–529 (2020).
65. Nieuwkoop, T., Claassens, N. J. & van der Oost, J. Improved protein production and codon optimization analyses in *Escherichia coli* by bicistronic design. *Microb. Biotechnol.* **12**, 173–179 (2018).
66. Rennig, M. et al. TARSyn: tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. *ACS Synth. Biol.* **7**, 432–442 (2018).
67. Arnold, F. H. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).
68. Copeland, N. G., Jenkins, N. A. & Court, D. L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769–779 (2001).
69. Bassalo, M. C. et al. Rapid and efficient one-step metabolic pathway integration in *E. coli*. *ACS Synth. Biol.* **5**, 561–568 (2016).
70. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
71. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206–223 (2010).
72. Gallagher, R. R., Li, Z., Lewis, A. O. & Isaacs, F. J. Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc.* **9**, 2301–2316 (2014).
73. Nyerges, Á. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl Acad. Sci. USA* **113**, 2502–2507 (2016).
74. Bonde, M. T. et al. Direct mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides. *ACS Synth. Biol.* **4**, 17–22 (2015).
75. Aparicio, T., Nyerges, A., Martínez-García, E. & de Lorenzo, V. High-efficiency multi-site genomic editing of *Pseudomonas putida* through thermoinducible ssDNA recombineering. *iScience* **23**, 100946 (2020).
76. Chang, Y., Wang, Q., Su, T. & Qi, Q. The efficiency for recombineering is dependent on the source of the phage recombinase function unit. Preprint at https://www.biorxiv.org/content/10.1101/745448v1.full.pdf (2019).
77. van Pijkeren, J. P., Neoh, K. M., Sirias, D., Findley, A. S. & Britton, R. A. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. *Bioengineered* https://doi.org/10.4161/bioe.2104 (2012).
78. Filsinger, G. T. et al. Characterizing the portability of phage-encoded homologous recombination proteins. *Nat. Chem. Biol.* **17**, 394–402 (2021).
79. Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl Acad. Sci. USA* **117**, 13689–13698 (2020).
This study provides a platform for identifying efficient single-stranded annealing proteins and allowed for more efficient recombineering in *E. coli* and several other bacteria.
80. Hueso-Gil, A., Nyerges, Á., Pál, C., Calles, B. & De Lorenzo, V. Multiple-site diversification of regulatory sequences enables interspecies operability of genetic eevices. *ACS Synth. Biol.* **9**, 104–114 (2020).
81. Barbieri, E. M., Muir, P., Akhuetie-Oni, B. O., Yellman, C. M. & Isaacs, F. J. Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes. *Cell* **171**, 1453–1467 (2017).
82. Dicarlo, J. E. et al. Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.* **2**, 741–749 (2013).
83. Liang, Z., Metzner, E. & Isaacs, F. J. Advanced eMAGE for highly efficient combinatorial editing of a stable genome. Preprint at https://www.biorxiv.org/content/10.1101/2020.08.30.256743v1 (2020).
84. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816–821 (2012).
85. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
86. Adiego-Pérez, B. et al. Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.* **366**, 1–19 (2019).
This review provides a comprehensive overview of CRISPR-Cas techniques and their performance for multiplexing in prokaryotic and eukaryotic microorganisms.
87. Zhang, Z. X. et al. Recent advances in the application of multiplex genome editing in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **105**, 3873–3882 (2021).
88. Kim, H., Ji, C., Je, H., Kim, J. & Kang, H. mpCRISTAR: multiple plasmid approach for CRISPR/Cas9 and TAR-mediated multiplexed refactoring of natural product biosynthetic gene clusters. *ACS Synth. Biol.* **9**, 175–180 (2020).

89. Lian, J., Schultz, C., Cao, M., HamediRad, M. & Zhao, H. Multi-functional genome-wide CRISPR system for high throughput genotype–phenotype mapping. *Nat. Commun.* **10**, 5794 (2019).
90. Napolitano, M. G. et al. Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **113**, E5588–E5597 (2016).
91. Ronda, C., Pedersen, L. E., Sommer, M. O. A. & Nielsen, A. T. CRMAGE: CRISPR optimized MAGE recombineering. *Sci. Rep.* **6**, 19452 (2016).
92. Oesterle, S., Gerngross, D., Schmitt, S., Roberts, T. M. & Panke, S. Efficient engineering of chromosomal ribosome binding site libraries in mismatch repair proficient *Escherichia coli*. *Sci. Rep.* **7**, 12327 (2017).
93. Asin-Garcia, E., Martin-Pascual, M., Garcia-Morales, L., Van Kranenburg, R. & Martins Dos Santos, V. A. P. ReScribe: an unrestrained tool combining multiplex recombineering and minimal-PAM ScCas9 for genome recoding *Pseudomonas putida*. *ACS Synth. Biol.* **10**, 2672–2688 (2021).
94. Jiang, Y. et al. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
95. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl. Environ. Microbiol.* **81**, 5103–5114 (2015).
96. Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2017).
Extensive work that established the CREATE platform for high-throughput targeted and trackable genome engineering in *E. coli* based on CRISPR-Cas9 and recombineering.
97. Liu, R. et al. Iterative genome editing of *Escherichia coli* for 3-hydroxypropionic acid production. *Metab. Eng.* **47**, 303–313 (2018).
98. Roy, K. R. et al. Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol.* **36**, 512–520 (2018).
99. Gaudelli, N. M. et al. Programmable base editing of A–T to G–C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
100. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
101. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
102. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).
103. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
104. Wang, Y. et al. In-situ generation of large numbers of genetic combinations for metabolic reprogramming via CRISPR-guided base editing. *Nat. Commun.* **12**, 678 (2021).
105. Volke, D. C., Martino, R. A., Kozaeva, E., Smania, A. M. & Nikel, P. I. Modular (de)construction of complex bacterial phenotypes by CRISPR/nCas9-assisted, multiplex cytidine base-editing. *Nat. Commun.* **13**, 3026 (2022).
This work established and demonstated an efficient base-editing tool for multiplex editing in the emerging model organism *P. putida*.
106. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T. & Lee, S. Y. A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.* **12**, 5206 (2021).
107. Reis, A. C. et al. Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat. Biotechnol.* **37**, 1294–1301 (2019).
108. Cui, L. et al. A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.* **9**, 1912 (2018).
109. Si, T., Luo, Y., Bao, Z. & Zhao, H. RNAi-assisted genome evolution in *Saccharomyces cerevisiae* for complex phenotype engineering. *ACS Synth. Biol.* **4**, 283–291 (2015).
110. Na, D. et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174 (2013).
111. Xie, W. H., Deng, H. K., Hou, J. & Wang, L. J. Synthetic small regulatory RNAs in microbial metabolic engineering. *Appl. Microbiol. Biotechnol.* **105**, 1–12 (2021).
112. Kiattisewee, C. et al. Portable bacterial CRISPR transcriptional activation enables metabolic engineering in *Pseudomonas putida*. *Metab. Eng.* **66**, 283–295 (2021).
113. Liu, Y., Wan, X. & Wang, B. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nat. Commun.* **10**, 3693 (2019).
114. Lin, J. L., Wagner, J. M. & Alper, H. S. Enabling tools for high-throughput detection of metabolites: metabolic engineering and directed evolution applications. *Biotechnol. Adv.* **35**, 950–970 (2017).
115. Leavell, M. D., Singh, A. H. & Kaufmann-Malaga, B. B. High-throughput screening for improved microbial cell factories, perspective and promise. *Curr. Opin. Biotechnol.* **62**, 22–28 (2020).
116. Sarnaik, A., Liu, A., Nielsen, D. & Varman, A. M. High-throughput screening for efficient microbial biotechnology. *Curr. Opin. Biotechnol.* **64**, 141–150 (2020).
117. Chen, L., Sun, G.-G., Fang, H.-T. & He, K. in *DEStech Transactions on Engineering and Technology Research* 110–117 (DEStech Publications, 2019); https://doi.org/10.12783/dtetr/amee2019/33444
118. Yang, H. et al. Study on fluorescence spectra of thiamine, riboflavin and pyridoxine. In *Proc. Seventh International Symposium on Precision Mechanical Measurements* 99030H (SPIE, 2016); https://doi.org/10.1117/12.2211248
119. Sesen, M. & Whyte, G. Image-based single cell sorting automation in droplet microfluidics. *Sci. Rep.* **10**, 8736 (2020).
120. Fu, L., Zhang, J. & Si, T. Recent advances in high-throughput mass spectrometry that accelerates enzyme engineering for biofuel research. *BMC Energy* **2**, 1 (2020).
121. Häbe, T. T. et al. Ultrahigh-throughput ESI-MS: sampling pushed to six samples per second by acoustic ejection mass spectrometry. *Anal. Chem.* **92**, 12242–12249 (2020).
122. Sinclair, I., Davies, G. & Semple, H. Acoustic mist ionization mass spectrometry (AMI-MS) as a drug discovery platform. *Expert Opin. Drug Discov.* **14**, 609–617 (2019).
123. Wyatt Shields Iv, C., Reyes, C. D. & López, G. P. Microfluidic cell sorting: a review of the advances in the separation of cells from debulking to rare cell isolation. *Lab Chip* **15**, 1230–1249 (2015).
124. Bouzetos, E., Ganar, K. A., Mastrobattista, E., Deshpande, S. & van der Oost, J. (R)evolution-on-a-chip. *Trends Biotechnol.* https://doi.org/10.1016/j.tibtech.2021.04.009 (2021).
125. Caen, O. et al. High-throughput multiplexed fluorescence-activated droplet sorting. *Microsyst. Nanoeng.* **4**, 33 (2018).
126. Guo, L., Zeng, W., Xu, S. & Zhou, J. Fluorescence-activated droplet sorting for enhanced pyruvic acid accumulation by *Candida glabrata*. *Bioresour. Technol.* **318**, 124258 (2020).
127. Liang, W. F. et al. Biosensor-assisted transcriptional regulator engineering for *Methylobacterium extorquens* AM1 to improve mevalonate synthesis by increasing the acetyl-CoA supply. *Metab. Eng.* **39**, 159–168 (2017).
This study nicely demonstrated the high-throughput targeting of a global regulator for improving mevalonate production by using an *in vivo* sensor for this product in a non-conventional, methylotrophic bacterium.
128. van Rossum, T., Kengen, S. W. M. & van der Oost, J. Reporter-based screening and selection of enzymes. *FEBS J.* **280**, 2979–2996 (2013).
129. Rogers, J. K. & Church, G. M. Genetically encoded sensors enable real-time observation of metabolite production. *Proc. Natl Acad. Sci. USA* **113**, 2388–2393 (2016).
130. Fang, M. et al. Intermediate-sensor assisted push-pull strategy and its application in heterologous deoxyviolacein production in *Escherichia coli*. *Metab. Eng.* **33**, 41–51 (2016).
131. Xu, P. et al. Design and kinetic analysis of a hybrid promoter-regulator system for malonyl-CoA sensing in *Escherichia coli*. *ACS Chem. Biol.* **9**, 451–458 (2014).
132. Tang, S.-Y. & Cirino, P. C. Design and application of a mevalonate-responsive regulatory protein. *Angew. Chem.* **123**, 1116–1118 (2011).
133. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* **42**, 84–91 (2016).
134. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl Acad. Sci. USA* **111**, 17803–17808 (2014).
This study demontrates the strenght of combining rational genome-wide targetting (MAGE) with product biosensors controlling selectable markers, leading to many-fold improved naringenin and glucaric acid production in *E. coli*.
135. Orsi, E., Claassens, N. J., Nikel, P. I. & Lindner, S. N. Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nat. Commun.* **12**, 5295 (2021).
136. Kim, S. et al. Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* **16**, 538–545 (2020).
137. Antonovsky, N. et al. Sugar synthesis from CO_2 in *Escherichia coli*. *Cell* **166**, 115–125 (2016).
138. Guzmán, G. I. et al. Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.* **15**, e8462 (2019).
139. Saleski, T. E. et al. Syntrophic co-culture amplification of production phenotype for high- throughput screening of microbial strain libraries. *Metab. Eng.* **54**, 232–243 (2019).
140. Patinios, C. et al. Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic Acids Res.* **49**, 11392–11404 (2021).
141. Chao, R., Mishra, S., Si, T. & Zhao, H. Engineering biological systems using automated biofoundries. *Metab. Eng.* **42**, 98–108 (2017).
142. Hillson, N. et al. Building a global alliance of biofoundries. *Nat. Commun.* **10**, 2040 (2019).

143. Isaacs, F. J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **6040**, 348–353 (2017).
144. Ostrov, N. et al. Design, synthesis and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
145. Szili, P. et al. Rapid evolution of reduced susceptibility against a balanced dual-targeting antibiotic through stepping-stone mutations. *Antimicrob. Agents Chemother.* **63**, e00207–e00219 (2019).
146. Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* **16**, 887–893 (2019).
147. Jung, S. W., Yeom, J., Park, J. S. & Yoo, S. M. Recent advances in tuning the expression and regulation of genes for constructing microbial cell factories. *Biotechnol. Adv.* **50**, 107767 (2021).

Acknowledgements

We thank R. van Kranenburg and E. Orsi for critical reading of this manuscript. S.Y., N.J.C. and J.v.d.O. acknowledge the support of the Dutch Research Council (NWO) via the Gravitation Project BaSyC (024.003.019) and Spinoza (SPI 93-537), awarded to J.v.d.O. In addition, N.J.C. acknowledges support from his NWO Veni fellowship (VI.Veni.192.156). Funding for this research was also provided by the US Department of Energy (DOE) under grant no. DE-FG02-02ER63445 and by the National Science Foundation (NSF) award no. 2123243 (both to G.M.C.). A.N. was supported by the EMBO LTF 160-2019 Long-Term fellowship.

Competing interests

J.v.d.O. is included as inventor on several CRISPR-related patents and is scientific advisor of NTrans Technologies, Scope Biosciences and Hudson River Biotechnology. G.M.C. is a founder of companies in which he has related financial interests: ReadCoor, EnEvolv and 64x Bio. For a complete list of G.M.C.'s financial interests, see also https://arep.med.harvard.edu/gmc/tech.html. A.N. is an inventor on a patent related to directed evolution with random genomic mutations (DIvERGE) (US patent 10669537B2: Mutagenizing intracellular nucleic acids). The remaining authors declare no competing interests.

Additional information

Correspondence should be addressed to Nico J. Claassens.

Peer review information *Nature Catalysis* thanks Sang Yup Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.