Minireview

Emerging Tools for Synthetic Genome Design

Bo-Rahm Lee[1], Suhyung Cho[1,2], Yoseb Song[1,2], Sun Chang Kim[1,2], and Byung-Kwan Cho[1,2,*]

Synthetic biology is an emerging discipline for designing and synthesizing predictable, measurable, controllable, and transformable biological systems. These newly designed biological systems have great potential for the development of cheaper drugs, green fuels, biodegradable plastics, and targeted cancer therapies over the coming years. Fortunately, our ability to quickly and accurately engineer biological systems that behave predictably has been dramatically expanded by significant advances in DNA-sequencing, DNA-synthesis, and DNA-editing technologies. Here, we review emerging technologies and methodologies in the field of building designed biological systems, and we discuss their future perspectives.

INTRODUCTION

With the remarkable development of sequencing technology, the genomes of diverse species from bacteria to human have been completely sequenced. The identification of unknown organisms *via* genome-sequencing analyses has also been rapidly performed as soon as the species are discovered (Pagani et al., 2012). In addition, sequencing speed has been greatly accelerated by the recent advent of next-generation sequencing (NGS) methods, which appeared in the last decade following the conventional, capillary-based sequencing methods. NGS methods analyze various sizes and types of DNA libraries more accurately, rapidly, and cheaply in a high-throughput fashion (Mardis, 2011; Metzker, 2010; Pareek et al., 2011; Shendure and Lieberman Aiden, 2012; Soon et al., 2013; Zhang et al., 2011). NGS provides powerful tools, not only to discover fundamental biological information, such as whole-genome sequences, variations, and DNA-protein interactions, but also to apply to diagnostics and clinical uses (Encode Project Consortium et al., 2012). For instance, the first human genome projects required 10 years and 3 years for sequencing and analysis, respectively; whereas current NGS technology allows the sequencing of multiple human genomes in a single run, requiring only a single week and reagent cost of less than $5,000 per genome (Lander et al., 2001; Soon et al., 2013; Venter et al., 2001).

The genome sequences contributed tremendously to all bioscience-related studies across biochemistry, agriculture, bioengineering, and medicine, as well as to the bioindustrial production of value-added products such as platform chemicals (Du et al., 2013; Esvelt and Wang, 2013; Heitzer et al., 2013; Reyes et al., 2012; Shokralla et al., 2012). However, the genome sequences did not merely make contributions to the deeper understanding of the molecular mechanisms of cellular function; they also provided engineering targets for various genome-manipulation technologies such as zinc-finger nuclease (ZFN) and transcription activator-like effector nucleases (TALENs) (Bedell et al., 2012a; Cong et al., 2013; Jiang et al., 2013; Kim et al., 2012a; Mali et al., 2013; Wang et al., 2012a). These recent breakthroughs in genome-engineering technologies are now offering new opportunities to produce rationally designed biological functions, which are critical issues that should be addressed by synthetic biology.

A major objective of synthetic biology is the design and engineering of biologically based parts, novel devices, and systems, as well as the redesign of existing biological systems (Endy, 2005; Nandagopal and Elowitz, 2011; Pleiss, 2006). A wide range of biological systems can be engineered to add new functions or to edit existing functions according to a user's purpose in a modular, reliable, and predictable way. This is possible because biological functions are inherently expressed through proteins and RNAs that are primarily encoded in DNA sequences. In addition, biological regulatory elements such as logic gates, feedback systems, and oscillators can be defined (Khalil and Collins, 2010). Thus, synthetic biology has great potential to deliver important new applications and improve existing industrial processes across many sectors including energy, pharmaceuticals, and materials (Martin et al., 2009; Medema et al., 2011; 2012). Here, we review the emerging tools currently available for synthetic biology that can be integrated to design and build novel biological systems. The future perspective of this new, emerging field is also discussed.

Genome analysis

Next-generation sequencing
DNA sequencing, which has been widely used over 30 years, has made great advances in molecular biology, genetics, diagnosis, and bioinformatics (Sanger et al., 1977). It had inevitable limitations, however, to generate whole-genome sequences by an individual researcher because of its speed, cost, and resolu-

Springer

tion. The advent of new technology for high-throughput DNA sequencing, NGS, overcame these barriers and caused a revolution in genome science.

Currently, three representative sequencers with different platform technologies are commercially available. The first commercial NGS platform was the Roche 454 System based on 'pyrosequencing'. Instead of sequencing in multiple tubes or microplate wells, the DNA library is amplified by emulsion polymerase chain reaction (PCR) on the surfaces of hundreds of thousands of beads (Dressman et al., 2003; Margulies et al., 2005). The current GS FLX+ system can sequence about 500 Mb of raw sequence in 23 h with a consensus accuracy of 99.99% and an average read length of 1,000 bp. Compared with other sequencing platforms, the Roche 454 system has an advantage in mapping repetitive regions because of its longer read length; but it has the disadvantage of an increasing error rate when sequencing a homopolymer: a row of repeats of the same short sequence (Rothberg and Leamon, 2008). The second commercial platform, the Illumina Genome Analyzer, is based on sequencing by synthesis and emerged in 2006. Different from the method using micro-beads mentioned above, the adaptor-linked DNA library binds to complementary adaptors immobilized in a flow cell and forms clusters via bridge amplification (Adessi et al., 2000). The clustered library produces a signal by incorporating four differentially labeled fluorescent dNTPs, and the sequencing progresses by repeating the process. The HiSeq 2,500 system currently produces a maximum read length of two 100 bp paired-end reads in high-output mode, which results in about 600 Gb of output within 11 days. The MiSeq system, a bench-top sequencer, produces a read length of two 250 bp paired-end reads and a maximum of 8.5 Gb of data per run. The third commercial system, the Sequencing by Oligonucleotide Ligation and Detection (SOLiD) sequencer using the Polonator technology, was commercially released in 2007 (Shendure et al., 2005). The open-source sequencer uses emulsion PCR, to immobilize the DNA library onto a solid support, and cyclic sequencing-by-ligation chemistry. A recent version, the 5,500 Series Genetic Analysis Systems, can sequence a maximum of 20 Gb of DNA per day. Personal sequencers, such as the MiSeq and Ion torrent™, of small size are efficient for relatively fast analyses. The difference of the Ion torrent™ system compared with other sequencing techniques is the use of a semiconductor-based sequencing technique rather than signal detection methods, such as the use of fluorescent dyes. In this system, when a dNTP is incorporated into DNA, it produces a pyrophosphate and a proton. The DNA library is immobilized in microwells, and when a dNTP is incorporated into the DNA by DNA polymerase, the change in pH caused by the protons can be detected as a voltage change. Sequencing data can be obtained rapidly and at low cost, because the Ion torrent™ system does not use expensive dNTPs labeled with fluorescent dyes and an optical detection device. The Ion PGM sequencer produces a maximum read length of 400 bp and generates between 1 and 2 Gb of data within several hours. The maximum read length and the total output of each platform is continuously increasing.

Unlike the NGS platforms that depend on DNA-library amplification, third-generation sequencing platforms determine DNA sequences directly from a single DNA molecule (Pareek et al., 2011; Schadt et al., 2010). This single-molecule sequencing can overcome sequencing errors and biases, caused by the favor of certain sequences during the amplification step, that potentially distort the abundances of various DNA fragments in the sample. These novel methods rely on sequencing by syn-

thesis, sequencing by degradation, or sequencing by direct physical inspection of the DNA molecule *via* a specially engineered DNA polymerase, various artificial nanopores, or advanced microscopy techniques (Clarke et al., 2009; Zhang et al., 2011). Although most third-generation sequencing platforms are currently still in development, they will offer advantages over current technologies, such as higher throughput, faster turnaround time, longer read lengths, higher accuracy, smaller amounts of starting material, and lower sequencing costs (Schadt et al., 2010).

Genome sequencing

With its high accuracy, fast speed, and low cost, NGS is widely used to determine the genome sequences of various organisms and the variations between genome sequences (Pushkarev et al., 2009; Wheeler et al., 2008). As of June 2012, a total of 3920 bacterial genomes and 854 eukaryotic genomes have been completely sequenced (Pagani et al., 2012). In addition to routine *de novo* genome sequencing, NGS can be also used for the diagnosis and monitoring of pathogens including viruses, bacteria, fungi, and parasites. NGS can produce full information about the genotype of a pathogen within several hours and help to formulate a drug for an unknown pathogen as soon as possible for public health purposes. NGS could replace complicated and time-consuming techniques used in routine clinical microbiology practices with a simple, more efficient workflow (Didelot et al., 2012; Shendure and Lieberman Aiden, 2012; Soon et al., 2013).

The obvious applications of NGS are *de novo* genome sequencing, whole-genome resequencing, and targeted resequencing. The first application of NGS to genome sequencing was to elucidate *Acinetobacter baumannii* pathogenesis using 454 sequencing technology (Smith et al., 2007). Despite the sequencing power of NGS, the short reads obtained from NGS draw a setback in terms of genome assembly and mapping applications (Wold and Myers, 2008). For example, repetitive sequences are widely distributed across the entire human genome, therefore the short reads are placed equally at multiple chromosomal locations. Efforts in bioinformatics have been made to overcome the limitation; but many challenges still exist to either efficiently assemble short reads *de novo* or else align them to a reference genome. The currently available bioinformatics tools for short-read alignment and *de novo* and reference-guided assembly have been reported elsewhere (Martin and Wang, 2011; Zhang et al., 2011).

Nevertheless, NGS allows for very deep genomic coverage. Whole-genome resequencing is therefore being widely performed for the correct identification of SNPs and structural variations such as insertions, deletions, copy number variations, and rearrangements. For example, a large number of human genomic variations have been determined by multi-institute consortia projects (1000 Genomes Project Consortium et al., 2010; Ball et al., 2012; Encode Project Consortitum, 2012; Hugo Pan-Asian SNP Consortium et al., 2009). To date, more than 30 million SNPs have been discovered by human genome sequencing projects. Copy number variation has also been shown to be associated with various diseases including glomerulone-phritis (Aitman et al., 2006) and Crohn's disease (McCarroll et al., 2008). Resequencing can also be applied to sequence an entire bacterial genome to identify acquired mutations. For instance, accumulated mutations in microbes adaptively evolved in the laboratory have been detected by comparing polymorphisms between ancestral and evolved genomes (Araya et al., 2010; Atsumi et al., 2010; Charusanti et al., 2010; Conrad et al.,

2009; Kishimoto et al., 2010).

Resequencing is not about analyzing the whole genome but instead sequencing subsets of genes or specific parts of genome. Targeted resequencing analyzes the interesting parts and concentrates studies on specific targets. Comparing the specific sites between different samples may assist the discovery of novel biological mechanisms (Qi et al., 2013a; Ram et al., 2011). Exome sequencing is one resequencing method that identifies protein-coding regions. The method is valuable for the diagnosis of genetic diseases, such as human pancreatic cancer (Wang et al., 2012b) and Mendelian disorders (Ku et al., 2011), caused by mutations in coding regions.

Transcriptome sequencing

The physiological properties of cells are programmed by a set of RNA molecules defined as the transcriptome. This means that distinct sets of genes are simultaneously activated or repressed by diverse regulators in accordance with changing environments. Thus, it is important to measure the levels of RNA transcripts to elucidate physiologically relevant biological processes. To readily analyze genome-wide cellular transcription, microarrays based on hybridization have been widely available (Schena et al., 1995; Selinger et al., 2000). However, the array-based approach is limited by a high rate of noise due to cross-hybridization, the dynamic range of detection due to signal saturation, and the inability to detect transcripts with low copy numbers per cell. In addition, array-based approaches require extensive normalization based on complicated statistical calculations to compare expression data from different experiments (Pinto et al., 2011).

NGS addressed the limitations of the array-based approaches with the introduction of RNA-seq, which was first invented in both yeast and mammalian cells (Cloonan et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Wilhelm et al., 2008). Unlike the array-based approaches, which depend on hybridization, RNA-seq allows the unambiguous mapping of transcripts to unique regions of the genome with single-base resolution. Hence, there is lower background noise (Soon et al., 2013). In addition to the accurate quantification of the transcriptome of known genes, RNA-seq allows the determination of correct gene annotation, expressed single nucleotide polymorphisms (SNPs), novel genes, and RNAs with high levels of reproducibility. Moreover, the strand-specific RNA sequencing principle is based on analyzing primary mRNA transcripts. Therefore, it allows the identification of more reliable and more accurate genome architectures (Cho et al., 2009; Kim et al., 2012b; Levin et al., 2010; Perkins et al., 2009; Qiu et al., 2010; Seo et al., 2012; Sharma et al., 2010).

Even without a reference genome, de novo transcriptome assembly can be done by using short reads from RNA-seq. This method has facilitated the reconstruction of the entire transcriptome (Grabherr et al., 2011; Martin and Wang, 2011). With the reconstructed transcriptome, BLAST or other gene prediction tools are used to identify functional annotation. Many assemblers, such as Multiple-k (Surget-Groba and Montoya-Burgos, 2010), Rnnotator (Martin et al., 2010), Trans-ABySS (Robertson et al., 2010), Oases (Schulz et al., 2012), and Trinity (Grabherr et al., 2011), have been developed. An example of de novo transcriptome assembly is the identification of novel genes related to useful biosynthetic pathways. Capsaicinoid from chili pepper, for example, is a practical compound that has various medical applications. Because the capsaicinoid biosynthetic pathway was not fully identified, de novo assembly was performed, and three novel structural genes in the biosynthetic pathway were discovered (Liu et al., 2013). Similarly, insecticide-resistance genes were identified in species of insect pests (Hsu et al., 2012). De novo transcriptome assembly can be more beneficial than genome sequencing alone, because RNA-seq not only reconstructs entire transcriptome, but it also measures the expression level of the target genes without genome sequences. Moreover, de novo transcriptome assembly with RNA-seq has the advantage of detecting additional transcription information, like alternative splicing sites in eukaryotes.

Interactome sequencing

To design and construct new synthetic circuits or pathways within a producer cell, synthetic biology demands the availability of standardized regulatory parts including promoters, ribosome-binding sites, terminators, DNA-binding proteins, and corresponding protein-binding sites (Cheng and Lu, 2012; Khalil and Collins, 2010; Nandagopal and Elowitz, 2011; Wang et al., 2013a). In this regard, NGS can be used to obtain the library of synthetic biology parts, enabling the development of novel devices and networks. The methods to identify the interactions between DNA and proteins (e.g., transcription factors) are chromatin immunoprecipitation (ChIP)-based techniques such as ChIP-seq and ChIP-exo (Johnson et al., 2007; Rhee and Pugh, 2011; Robertson et al., 2007). For ChIP-seq, DNA-protein complexes are specifically isolated by an antibody against the target protein. Purified DNA obtained from the immunoprecipitated DNA-protein complexes is ligated with sequencing adaptors, amplified by PCR, and massively sequenced. Unlike electrophoretic mobility shift assay, the sequencing results verify in vivo interactions between proteins and their corresponding binding sites. Although microarrays have been frequently used to map DNA-protein or RNA-protein interactions at the genome scale, NGS technologies are quickly replacing the use of microarrays. Compared with previous microarray-based results, ChIP-seq has superior resolution, requires less input DNA, produces less background noises, and has a better detection limit (Furey, 2012; Park, 2009; Valouev et al., 2008). In addition, NGS can be used to determine the interaction between RNA-binding proteins and RNA via a technique called ultraviolet (UV) crosslinking and immunoprecipitation (CLIP). After the irradiation of cells with UV 254 nm (HITS-CLIP) (Yeo et al., 2009) or UV 365 nm (PAR-CLIP) (Hafner et al., 2010), the cells are lysed, and RNA-protein complexes are pulled down using antibody-immobilized beads. Like ChIP-seq, protein-binding sites on RNAs are collected from the sequencing library.

ChIP-seq determines the targets of transcription factors in bacteria and eukaryotes. For instance, GlxR is one of the key transcriptional regulators in Corynebacterium glutamicum, an amino acid-producing bacterium. The investigation of C. glutamicum's regulatory network using ChIP-seq revealed 21 novel binding sites and concluded that GlxR is responsible for carbon-source metabolism and energy conversion (Jungwirth et al., 2013). Taken together, this approach will greatly expand our understanding of genome-wide associations between proteins and DNA. It will also grow the library of synthetic biology parts, thus enhancing our ability to model cellular behavior.

Genome design

Minimal genome as a programmable biological chassis

Approximately 20% of the 4,000 genes found in E. coli have not been functionally annotated (Keseler et al., 2013). In addition, numerous other cellular components with unknown genes are
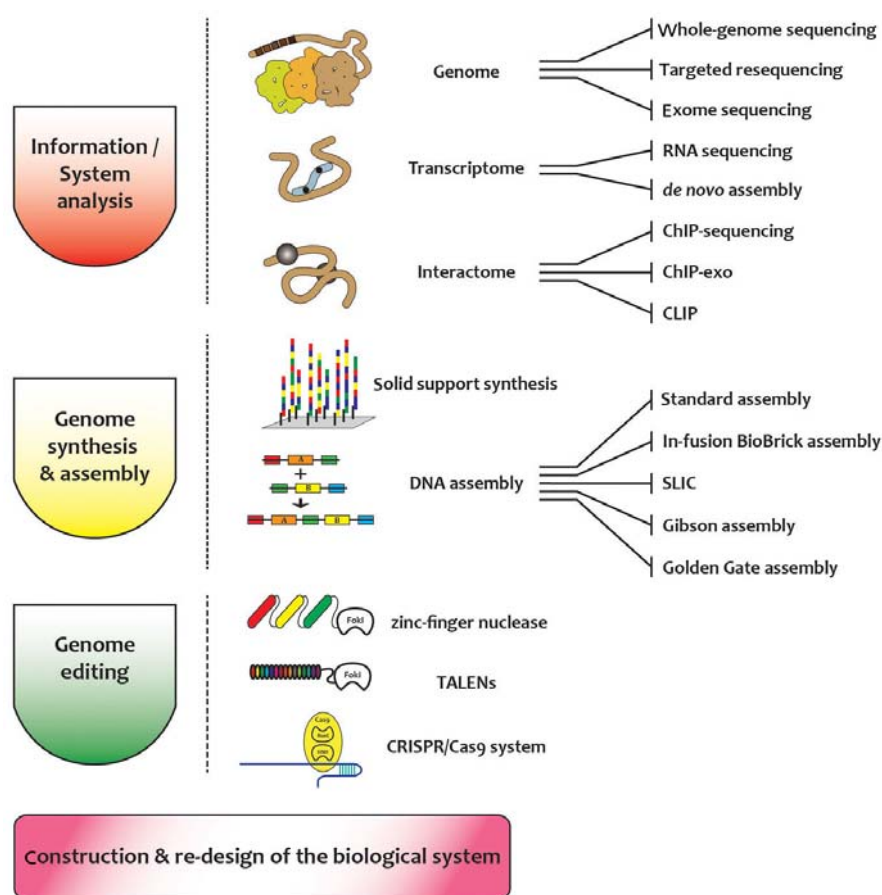
Fig. 1. Overall scheme and related synthetic biology tools for building new biological systems.

the most significant barriers to rational genome design. The genome-reduction approach provides a chassis for the construction of synthetic genomes with improved metabolic efficiency by minimizing redundant genes and regulatory circuits. Four criteria can guide the design of a useful biological chassis: growth rate, fermentation ability, genetic manipulation, and additional gene deletion. For example, the deletion of large blocks of nonessential genes that are not needed for the metabolic pathways of interest can reduce the production of unwanted by-products, increase the genome stability, and simplify the metabolism without physiological compromise (Forster and Church, 2006; Giga-Hama et al., 2007; Glass et al., 2006; Morimoto et al., 2008; Posfai et al., 2006; Yu et al., 2002).

A method for creating random and repeated genomic deletions was developed to create deletions in *Escherichia coli* MG1655 using a Tn5 derivative: essentially, a rearranged composite transposon that promotes the use of the internal transposon ends (Goryshin et al., 2003). The method was used to minimize the *E. coli* MG1655 genome down to an average of 200 kb (averaging 10 kb per deletion) through 20 cycles of random integration and deletion steps. The deletion locations were confirmed by microarray-hybridization experiments. Lee *et al.* reengineered the *E. coli* MDS42 chromosome to be 14.3% smaller than that of its parental *E. coli* strain, MG1655 (Lee et al., 2009). The genome reduction increased the production of L-threonine in MDS42 by overexpressing a feedback-resistant threonine operon (*thrA*BC*), deleting the genes that encode

threonine dehydrogenase (*tdh*) and threonine transporters (*tdcC* and *sstT*), and introducing a mutant threonine exporter (*rhtA23*) (Lee et al., 2009). The resulting strain, MDS-205, showed an ~83% increase in L-threonine production compared with a wild-type *E. coli* strain. Westers et al. (2003) deleted 0.53 Mb (7.7% of the genome) from *Bacillus subtilis*, whose wild-type genome size is 4.2 Mb. Another *B. subtilis* strain, MGB874, was created by step-by-step deletions of 28 regions (20.7% of the genome) in which single deletions do not affect cell growth (Morimoto et al., 2008). When plasmids encoding extracellular cellulase and protease enzymes were transformed into MGB874, their productivities were enhanced up to 1.7- and 2.5-fold, respectively, relative to the wild-type strain. Genome reduction in *Saccharomyces cerevisiae* was also attempted to create a mutant genome in which 531.5 kb of the wild-type genome was deleted (Murakami et al., 2007). This mutant increased its ethanol and glycerol production by 1.8- and 2-fold, respectively; and its resistance to external stresses such as temperature, pH, salt, and changes of growth media was equivalent to that of the wild type. The deletion of a large genomic region in the fission yeast *Saccharomyces pombe* to find an efficient mutant for heterologous protein production was also achieved using homologous recombination between the chromosome and a fragment of linear DNA (Giga-Hama et al., 2007; Hirashima et al., 2006).

The genome-reduction approach has been focused on the deletion of unnecessary genes. Because the goal of synthetic biology is to implement in a cell artificial regulatory circuits that
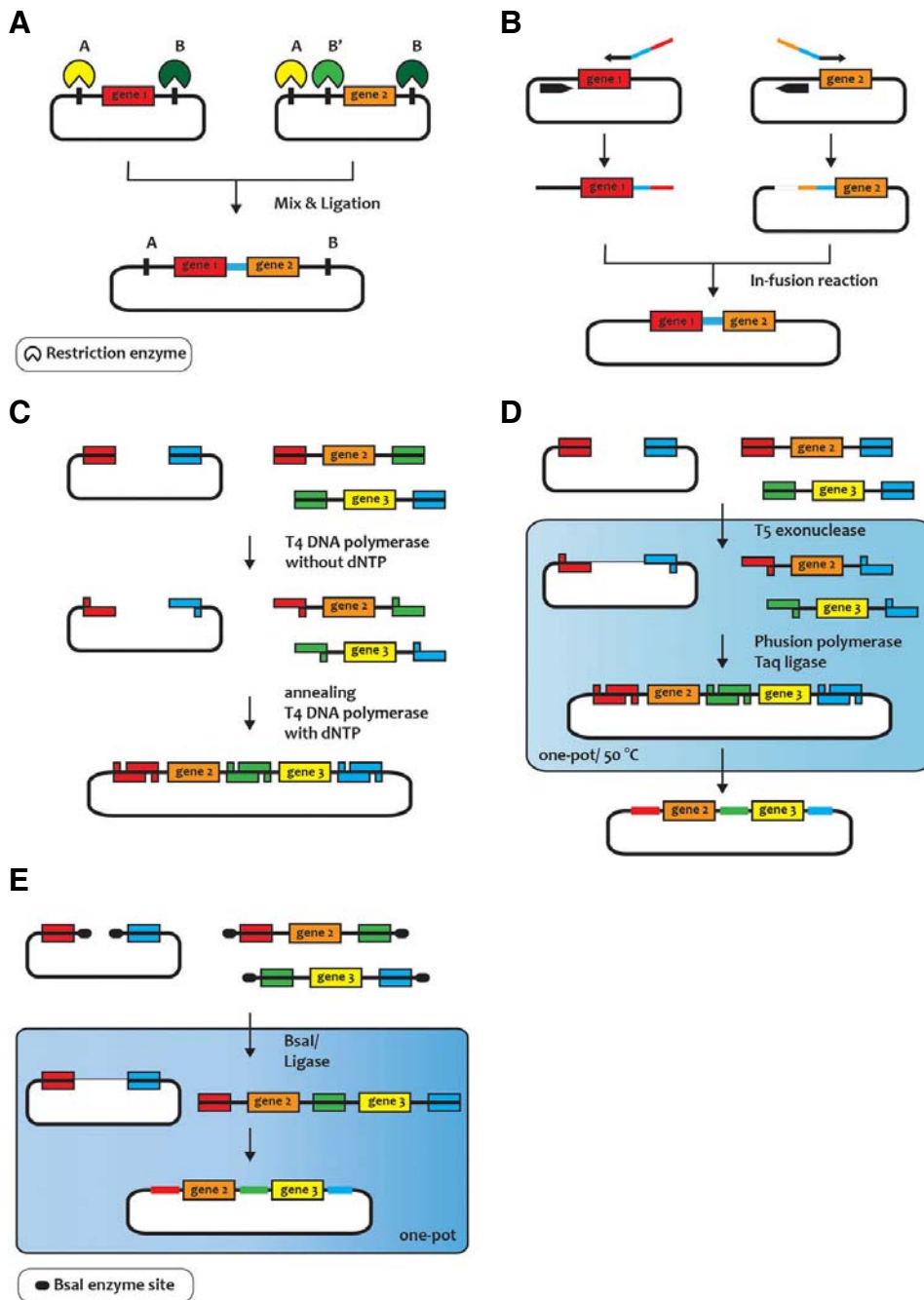
can be predictable, measurable, and controllable, the vast number of endogenous regulatory interactions should reduced and replaced by well-defined orthogonal equivalents.

Genome synthesis and assembly

The first synthetic genome was that of poliovirus (Cello et al., 2002). Subsets of the full-length viral genome were first synthesized and cloned separately into plasmids that were subsequently used to transform living cells. Recently, Gibson et al. (2010) achieved whole-genome design, synthesis, and assem-

bly. They synthesized the 1.08 Mb *Mycoplasma mycoides* JCVI-syn1.0 genome starting from digitized genome-sequence information. Furthermore, they transplanted the chemically synthesized genome into a *Mycoplasma capricolum* recipient cell to make variants of *M. mycoides* that are controlled only by the synthetic chromosome. Thus, genome engineering is being greatly advanced by the improvement of DNA synthesis techniques in rapid time at an affordable price. The synthesis capacity has broadened from producing small primers to *de novo* synthesis of genes with any length and any complexity. Gene-scale synthesis is also becoming highly commercialized.
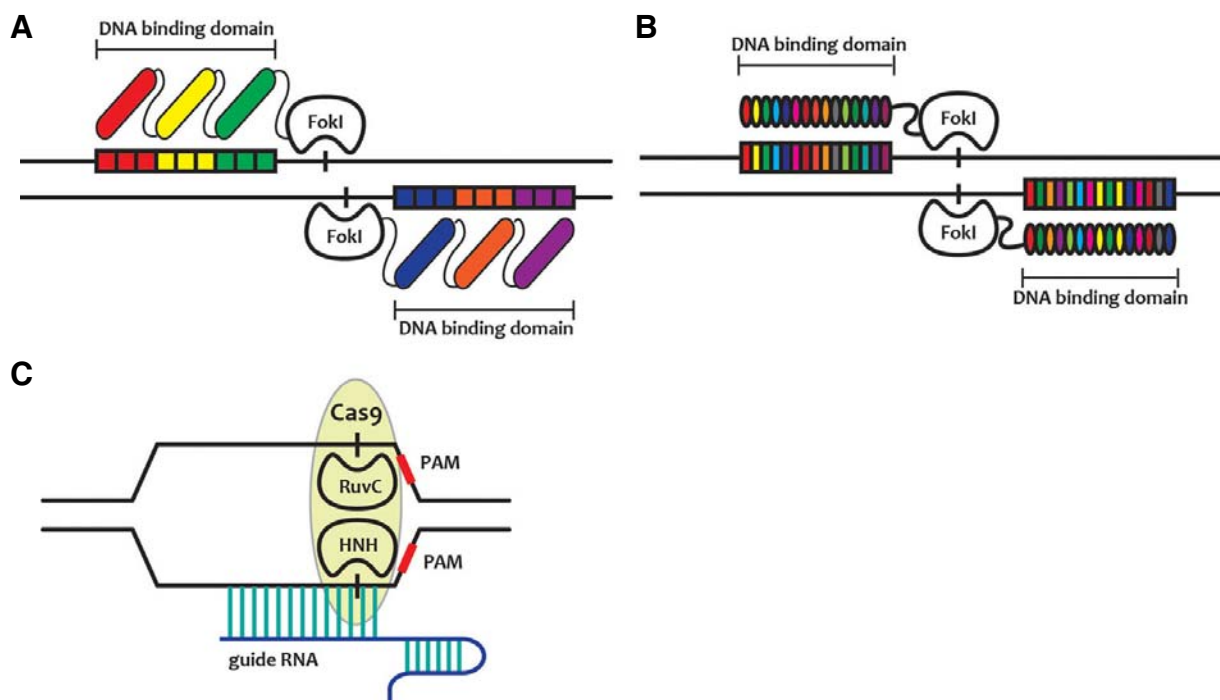
Fig. 3. Schematic diagram of genome editing methods. (A) Zinc-finger nuclease (ZFNs), which is recognized and bound with three nucleotides of the genome; (B) Transcription activator-like effector nuclease (TALENs), in which each amino acid in the DNA-binding domain recognizes one nucleotide; and (C) Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system, which uses guide RNAs for recognizing target DNA.

It is now inexpensive: near $0.29 per bp. The chemical synthesis of up to 4.5 Mb in a month is now feasible. The high-quality error-correction methods save time and labor at a reasonable price, frequently requiring less time and labor than conventional molecular cloning and sequencing (Richmond et al., 2004; Tian et al., 2004).

Although the accuracy of the traditional DNA synthesis method is very high, the quantity of each reaction batch limits the synthesis of whole genomes. To overcome this problem, attempts have been made to synthesize massive numbers of oligonucleotides on solid supports, like beads and microarrays, in a high-throughput fashion (Tian et al., 2004). In 2010, the Church group introduced a new technology using a combinatorial method of microarray and Roche 454 to synthesize a target DNA library. First, microarray-derived oligonucleotide pools were amplified and sequenced using Roche 454. After the sequencing, the correctly synthesized oligonucleotide was extracted, and the DNA was re-amplified. Compared with the microarray library, the finalized library-error rate was reduced by 500 fold (Matzas et al., 2010). Many studies on using microarray-to-gene synthesis and expanding the length of the oligonucleotide have been attempted. The microarray-derived oligonucleotides were selectively amplified by specific short primers. Using the amplified oligonucleotide pool, the target genes were assembled, producing up to a total of 35 kb (Kosuri et al., 2010). In addition, a 0.5 to 1 kb oligonucleotide was individually synthesized on subarrays and endonucleased by the mismatch-specific endonuclease. This approach reduced the error rate to 0.19 errors per kb and assembled the target gene (Quan et al., 2011).

Gene assembly methods have been used to build targeted

genetic circuits. For example, the BioBricks library contains various promoters, ribosome-binding sites (RBSs), coding sequences, and transcriptional terminators (Canton et al., 2008). The standard assembly, as shown in Fig. 2A, uses restriction enzymes. Using restriction enzyme and ligase, the target genes and the destination vector are given the same two sticky ends and ligated to form a new plasmid. However, this method requires different numbers of restriction-enzyme sites, depending on the number of target genes. After the ligation, a scar sequence may remain because of the effects of the restriction sites. The digested products also need time-consuming purification steps. An in-fusion BioBrick assembly (Fig. 2B) was newly developed to avoid the purification steps and the scar sequence (Sleight et al., 2010). The sequence and ligase-independent cloning (Fig. 2C) method was developed to avoid the use of restriction and ligation enzymes (Li and Elledge, 2007). Using PCR, the target gene with 25-bp sequence homologous to the linearized destination vector is amplified. Without dNTPs, the 3′ exonuclease activity of T4 DNA polymerase makes 5′ overhang. After the exonuclease activity, the annealing and adding of dNTPs trigger the homologous recombination of the target gene to form a new vector. Gibson assembly (Fig. 2D) uses a linearized destination vector and PCR products and the 5′ exonuclease activity of T5 exonuclease (Gibson et al., 2009). Golden Gate assembly (Fig. 2E), another method of DNA assembly (Engler et al., 2008) that uses type IIS endonuclease, has different recognition and cleavage sites that produce insert-DNA and a vector. This simple procedure helps multiple target assemblies to be worked simultaneously. The circular polymerase-extension cloning method, which uses double-stranded DNA and a linearized destination vector, was de-

veloped to assemble more efficiently (Quan and Tian, 2009; Zhang et al., 2012). The uracil excision-based cloning method solved the difficulty in controlling the exonuclease activity (Bitinaite et al., 2007) and the DNA-fragment assembly by introducing nicking endonuclease (Wang et al., 2013b).

Genome editing tools

An in-depth understanding of the complex genetic functions of complete genome sequences requires extensive genetic, biochemical, cytological, and physiological analyses. Likewise, the design of new biological systems demands the precise manipulation of genome sequences. In this context, genome editing is a powerful technology that introduces predetermined sequence changes to specific genomic loci (Perez-Pinera et al., 2012). This has been facilitated by the development of engineered nucleases with versatile and programmable DNA-binding domains, such as zinc-finger proteins and transcription activator-like effectors that are fused to the catalytic domain of a restriction endonuclease (Esvelt and Wang, 2013; Perez-Pinera et al., 2012). These artificially engineered nucleases are used to create specific double-stranded breaks (DSBs) at desired locations in the genome, which are subsequently repaired by endogenous homology-directed repair (HDR) pathways or nonhomologous end-joining (NHEJ) (Brugmans et al., 2007; Kanaar et al., 1998; Sung and Klein, 2006). HDR is an error-free DSB repair system that uses homologous DNA as a template. NHEJ often induces small insertions and deletions at breakpoint junctions, because it uses non-homologous template DNA (Gaj et al., 2012). Currently, two families of engineered nucleases are commercially available: ZFN and TALENs.

Zinc finger nucleases

ZFNs are chimeric endonucleases generated by the fusion of the nonspecific cleavage domain of the type IIS FokI restriction endonuclease with custom-designed Cys2-His2-zinc-finger proteins (Kim and Chandrasegaran, 1994). ZFNs have been used as a powerful genome-editing tools to generate targeted mutagenesis (Bibikova et al., 2002; Maeder et al., 2008; Urnov et al., 2005) and chromosomal rearrangements (Brunet et al., 2009; Lee et al., 2010) in numerous species including *Caenorhabditis elegans* (Morton et al., 2006), *Drosophila melanogaster* (Bibikova et al., 2002), silkworms (Takasu et al., 2010), monarch butterflies (Merlin et al., 2012), zebrafish (Doyon et al., 2008; Meng et al., 2008), sea urchins (Ochiai et al., 2010), *Arabidopsis thaliana* (Osakabe et al., 2010; Zhang et al., 2010), tobacco (Cai et al., 2009; Maeder et al., 2008; Townsend et al., 2009), corn (Shukla et al., 2009), mice (Carbery et al., 2010; Meyer et al., 2010), rats (Geurts et al., 2009), and humans (Hockemeyer et al., 2009; Holt et al., 2010; Zou et al., 2009).

A single-stranded breaks system, a modification of the tools that use ZFNs to induce DSBs, was also developed by fusing DNA-nicking enzymes (nickases) with zinc-finger proteins. It allowed accurate, site-specific genome modifications at only the on-target site (Brunet et al., 2009; Kim et al., 2012a). In addition, a ZFNickase that stimulates HDR but not NHEJ was developed by the mutation of a critical residue for FokI cleavage activity in one monomer of the ZFN heterodimer. It allowed the ZFNs to heterodimerize on DNA, but it restricted the cleavage to a single DNA strand. Consequently, ZFNickase induced gene addition at an endogenous CCR5 locus without the significant increase caused by the error-prone NHEJ repair pathway (Wang et al., 2012a).

Transcription activator-like effector nucleases

Transcription activator-like (TAL) effectors expressed by bacterial plant pathogens of the genus *Xanthomona Xanthomonas* consist of the translocation domain at the N-terminus, the repeated domain for DNA binding, nuclear localization, and the transcriptional activation domain at the C-terminus. The TAL effector repeats are highly conserved domains consisting of 33 to 35 amino acids, except for two internal positions known as repeat variable di-residues that display specificity to a target gene (Boch et al., 2009; Moscou and Bogdanove, 2009). By means of this specific protein, a TAL effector was engineered to produce a TALEN (Bogdanove and Voytas, 2011; Carlson et al., 2012) that associates with the FokI nuclease, which has an N-terminal DNA-binding domain and a non-specific DNA cleavage domain at the C-terminus (Wah et al., 1997). To function as a dimer, the FokI domain requires two constructs of unique DNA-binding domains with proper orientation and spacing in the target genome. Therefore, TALENs can be used to generate site-specific DSBs that can be repaired by joining, either by NHEJ or by homology-directed repair without donor DNAs, resulting in gene insertion or correction as well as the disruption of the gene of interest in various species such as yeast (Li et al., 2011), Xenopus embryos (Lei et al., 2012), plants (Cermak et al., 2011), nematodes (Wood et al., 2011), zebrafish (Huang et al., 2011; Sander et al., 2011), rats (Tesson et al., 2011), and human somatic (Cermak et al., 2011; Matsumura et al., 2010; Miller et al., 2011) and pluripotent stem cells (Hockemeyer et al., 2011).

The TALENs are broadly applied to custom-designed genome-editing tools in a high-throughput manner (Miller et al., 2011; Reyon et al., 2012). Reyon *et al.* designed the fast ligation-based automatable solid-phase high-throughput (FLASH) system, which is fast and economical for the large-scale assembly of TALENs (Reyon et al., 2012). Recently, Bedell VM *et al.* provided an advanced TALENs approach for targeted zebrafish genome editing, which was a bottleneck in the genome modification field (Bedell et al., 2012b).

RNA-guided CRISPR nucleases

Clustered regularly interspaced short palindromic repeats (CRISPRs) are short, repetitive genomic sequences that function as an immune system in 90% of reported archaea and 40% of known bacteria (Jansen et al., 2002; Makarova et al., 2011; Sorek et al., 2008). The CRISPR-associated system cleaves viral and plasmid DNAs into short fragments, which are integrated into the host chromosome at one end of the repeated region of the CRISPR locus. When the foreign DNA is transferred into the host cell, the records of prior integration events are transcribed and used to produce a library of short CRISPR-derived RNAs (crRNAs) by a CRISPR-associated or RNase III family nuclease. Each crRNA is complementary to a foreign target DNA and mediates the sequence-specific cleavage of the foreign DNA by the endonuclease function of Cas proteins (Makarova et al., 2011; Sorek et al., 2008; Wiedenheft et al., 2012). For example, the Type II CRISPR system from *Streptococcus pyogenes* consists of a single gene encoding the Cas9 protein, a mature crRNA, and a partially complementary trans-acting RNA (tracrRNA) (Jinek et al., 2012). In this system, a single CRISPR-associated protein (Cas9) in complex with the crRNA and the tracrRNA catalyzes the cleavage of complementary sequences, called proto-spacers, present within the target DNA. Upon recognizing the 20-nucleotide sequence, guided by the crRNA, Cas9 generates blunt DSBs at the site 3 bp upstream of the conserved proto-spacer-adjacent motif

(PAM) (Jinek et al., 2012; Mali et al., 2013). Subsequently, the DSBs stimulate NHEJ repair in genomic DNA. Although the PAMs are short DNA sequences (NGG, NGGNG, NNAGAAW, and NAAR), they are necessary, in addition to at least 12 bp of perfect homology, for CRISPR endonuclease activity (Cong et al., 2013; Deltcheva et al., 2011; Deveau et al., 2008; Esvelt and Wang, 2013; Horvath et al., 2008; Jinek et al., 2012; Mali et al., 2013; van der Ploeg, 2009). The requirement of tracrRNA, which plays an important role in the maturation of crRNA (Deltcheva et al., 2011; Horvath et al., 2008), can be engineered to a designed, hairpin-structured single-guide RNA (sgRNA) that mimics the tracrRNA-crRNA complex (Jinek et al., 2012). Thus, the engineered CRISPR system requires the Cas9 protein and the sgRNA, and therefore represents a facile strategy for targeted genome editing. Recently, it has been demonstrated that the engineered CRISPR system can be applied to genome editing in humans (Cho et al., 2013; Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013), zebrafish embryos (Hwang et al., 2013), *S. cerevisiae* (Dicarlo et al., 2013), and bacterial cells (Jiang et al., 2013; Jinek et al., 2012), as well as to the RNA-guided programming of gene expression (Hale et al., 2012; Qi et al., 2012; 2013b).

CONCLUSION

The ultimate goal of synthetic biology is to build novel biological systems that have new functions or to engineer existing biological systems to have better efficiency. For this, it is useful to consider the four core technical areas of synthetic biology: indexing biological parts, DNA synthesis and assembly, sequencing, and genome editing. Indeed, NGS platforms have addressed the technical barriers to securing enormous numbers of biological parts, such as promoters, RBSs, terminators, and regulatory protein-binding sites, as well as new genes including enzymes and regulatory proteins. Because the functions of biological parts are host-specific in many cases, some may not function as expected in their new context. Thus, host-specific libraries of biological parts, such as the strength of RBSs or regulatory sequences, will be useful to build new biological systems. In addition, the desired biological parts are often not single genes but are instead complete operons, such as a biosynthetic pathway consisting of several genes. Furthermore, the candidate parts need to be combined into transcriptional units with a well-designed regulatory circuitry in the context of genome-scale metabolic and regulatory networks. Manually curated, genome-scale biochemical reconstructions of around 40 different bacteria and the community-driven, global reconstruction of the human metabolism will be useful to predict the suitability of the transcriptional unit (Feist et al., 2009; Thiele et al., 2013).

Although the decrease in the cost of DNA synthesis has dramatically accelerated the construction of thousands of variants of DNA parts, codon optimization is also important for the efficient heterologous expression of the desired biological parts. For this, several computational algorithms are currently available. In addition, the fast and efficient formation of new gene clusters requires the simultaneous assembly of biological parts in a robust and flexible manner. New genome editing methods are also bringing down the number of errors in the genomes synthesized with the current sequencing capabilities. In addition to these experimental efforts, it is also crucial to develop computational algorithms for the *de novo* design of metabolic pathways with regulatory circuitry. With the many challenges to the understanding of natural biological systems, the rapid progress of emerging tools for synthetic biology has begun to provide genomes for applications in the areas of energy, health care, biochemicals, and the environment.

REFERENCES

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.J., Mayer, P., and Kawashima, E. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res. *28*, E87.

Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E., et al. (2006). Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature *439*, 851-855.

Araya, C.L., Payen, C., Dunham, M.J., and Fields, S. (2010). Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics *11*, 88.

Atsumi, S., Wu, T.Y., Machado, I.M., Huang, W.C., Chen, P.Y., Pellegrini, M., and Liao, J.C. (2010). Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. Mol. Syst. Biol. *6*, 449.

Ball, M.P., Thakuria, J.V., Zaranek, A.W., Clegg, T., Rosenbaum, A.M., Wu, X., Angrist, M., Bhak, J., Bobe, J., Callow, M.J., et al. (2012). A public resource facilitating clinical use of genomes. Proc. Natl. Acad. Sci. USA *109*, 11920-11927.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y., et al. (2012a). *In vivo* genome editing using a high-efficiency TALEN system. Nature *491*, 114-118.

Bedell, V.M., Person, A.D., Larson, J.D., McLoon, A., Balciunas, D., Clark, K.J., Neff, K.I., Nelson, K.E., Bill, B.R., Schimmenti, L.A., et al. (2012b). The lineage-specific gene ponzr1 is essential for zebrafish pronephric and pharyngeal arch development. Development *139*, 793-804.

Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics *161*, 1169-1175.

Bitinaite, J., Rubino, M., Varma, K.H., Schildkraut, I., Vaisvila, R., and Vaiskunaite, R. (2007). USER friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res. *35*, 1992-2002.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. Science *333*, 1843-1846.

Brugmans, L., Kanaar, R., and Essers, J. (2007). Analysis of DNA double-strand break repair pathways in mice. Mut. Res. *614*, 95-108.

Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. Proc. Natl. Acad. Sci. USA *106*, 10620-10625.

Cai, C.Q., Doyon, Y., Ainley, W.M., Miller, J.C., Dekelver, R.C., Moehle, E.A., Rock, J.M., Lee, Y.L., Garrison, R., Schulenberg, L., et al. (2009). Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol. Biol. *69*, 699-709.

Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. Nat. Bio-

technol. *26*, 787-793.

Carbery, I.D., Ji, D., Harrington, A., Brown, V., Weinstein, E.J., Liaw, L., and Cui, X. (2010). Targeted genome modification in mice using zinc-finger nucleases. Genetics *186*, 451-459.

Carlson, D.F., Fahrenkrug, S.C., and Hackett, P.B. (2012). Targeting DNA with fingers and TALENs. Mol. Ther. Nucleic Acids *1*, e3.

Cello, J., Paul, A.V., and Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science *297*, 1016-1018.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. *39*, e82.

Charusanti, P., Conrad, T.M., Knight, E.M., and Palsson, B. (2010). Genetic Basis of Growth adaptation of E. coli after deletion of pgi, a major metabolic gene. PLoS Genet. *6*, e1001186.

Cheng, A.A., and Lu, T.K. (2012). Synthetic biology: an emerging engineering discipline. Annu. Rev. Biomed. Eng. *14*, 155-178.

Cho, B.K., Zengler, K., Qiu, Y., Park, Y.S., Knight, E.M., Barrett, C.L., Gao, Y., and Palsson, B.O. (2009). The transcription unit architecture of the Escherichia coli genome. Nat. Biotechnol. *27*, 1043-1049.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230-232.

Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. *4*, 265-270.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat. Methods *5*, 613-619.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY) *339*, 819-823.

Conrad, T.M., Joyce, A.R., Applebee, M.K., Barrett, C.L., Xie, B., Gao, Y., and Palsson, B.O. (2009). Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. *10*, R118.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Strep-tococcus thermophilus. J. Bacteriol. *190*, 1390-1400.

Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. *41*, 4336-4343.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E., and Crook, D.W. (2012). Transforming clinical microbiology with bacterial genome sequencing. Nat. Rev. Genet. *13*, 601-612.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., et al. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat. Biotechnol. *26*, 702-708.

Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc. Natl. Acad. Sci. USA *100*, 8817-8822.

Du, R., Mercante, D., and Fang, Z. (2013). An artificial functional family filter in homolog searching in next-generation sequencing metagenomics. PLoS One *8*, e58669.

Encode Project Consortium, Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Endy, D. (2005). Foundations for engineering biology. Nature *438*, 449-453.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

Esvelt, K.M., and Wang, H.H. (2013). Genome-scale engineering for systems and synthetic biology. Mol. Syst. Biol. *9*, 641.

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B.O. (2009). Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. *7*, 129-143.

Forster, A.C., and Church, G.M. (2006). Towards synthesis of a minimal cell. Mol. Syst. Biol. *2*, 45.

Furey, T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat. Rev. Genet. *13*, 840-852.

Gaj, T., Guo, J., Kato, Y., Sirk, S.J., and Barbas, C.F., 3rd (2012). Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat. Methods *9*, 805-807.

Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., Jenkins, S.S., Wood, A., Cui, X., Meng, X., et al. (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. Science *325*, 433.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343-345.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science *329*, 52-56.

Giga-Hama, Y., Tohda, H., Takegawa, K., and Kumagai, H. (2007). Schizosaccharomyces pombe minimum genome factory. Biotechnol. Appl. Biochem. *46*, 147-155.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., 3rd, Smith, H.O., and Venter, J.C. (2006). Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. USA *103*, 425-430.

Goryshin, I.Y., Naumann, T.A., Apodaca, J., and Reznikoff, W.S. (2003). Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis. Genome Res. *13*, 644-653.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644-652.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., et al. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol. Cell *45*, 292-302.

Heitzer, E., Auer, M., Gasch, C., Pichler, M., Ulz, P., Hoffmann, E.M., Lax, S., Waldispuehl-Geigl, J., Mauermann, O., Lackner, C., et al. (2013). Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res. [Epub ahead of print].

Hirashima, K., Iwaki, T., Takegawa, K., Giga-Hama, Y., and Tohda, H. (2006). A simple and effective chromosome modification method for large-scale deletion of genome sequences and identification of essential genes in fission yeast. Nucleic Acids Res. *34*, e11.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat. Biotechnol. *27*, 851-857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. *29*, 731-734.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C., et al. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 *in vivo*. Nat. Biotechnol. *28*, 839-847.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barran-

gou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. *190*, 1401-1412.

Hsu, J.C., Chien, T.Y., Hu, C.C., Chen, M.J., Wu, W.J., Feng, H.T., Haymer, D.S., and Chen, C.Y. (2012). Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a de novo assembled transcriptome. PLoS One *7*, e40950.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat. Biotechnol. *29*, 699-700.

Hugo Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.H., Chen, J., et al. (2009). Mapping human genetic diversity in Asia. Science (New York, N.Y.) *326*, 1541-1545.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227-229.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. *43*, 1565-1575.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. ELife *2*, e00471.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of *in vivo* protein-DNA interactions. Science *316*, 1497-1502.

Jungwirth, B., Sala, C., Kohl, T.A., Uplekar, S., Baumbach, J., Cole, S.T., Puhler, A., and Tauch, A. (2013). High-resolution detection of DNA binding sites of the global transcriptional regulator GlxR in Corynebacterium glutamicum. Microbiology *159*, 12-22.

Kanaar, R., Hoeijmakers, J.H., and van Gent, D.C. (1998). Molecular mechanisms of DNA double strand break repair. Trends Cell Biol. *8*, 483-489.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2013). EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. *41*, D605-612.

Khalil, A.S., and Collins, J.J. (2010). Synthetic biology: applications come of age. Nat. Rev. Genet. *11*, 367-379.

Kim, Y.G., and Chandrasegaran, S. (1994). Chimeric restriction endonuclease. Proc. Natl. Acad. Sci. USA *91*, 883-887.

Kim, E., Kim, S., Kim, D.H., Choi, B.S., Choi, I.Y., and Kim, J.S. (2012a). Precision genome engineering with programmable DNA-nicking enzymes. Genome Res. *22*, 1327-1333.

Kim, D., Hong, J.S., Qiu, Y., Nagarajan, H., Seo, J.H., Cho, B.K., Tsai, S.F., and Palsson, B.O. (2012b). Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. PLoS Genet. *8*, e1002867.

Kishimoto, T., Iijima, L., Tatsumi, M., Ono, N., Oyake, A., Hashimoto, T., Matsuo, M., Okubo, M., Suzuki, S., Mori, K., et al. (2010). Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. PLoS Genet. *6*, e1001164.

Kosuri, S., Eroshenko, N., Leproust, E.M., Super, M., Way, J., Li, J.B., and Church, G.M. (2010). Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat. Biotechnol. *28*, 1295-1299.

Ku, C.S., Naidoo, N., and Pawitan, Y. (2011). Revisiting Mendelian disorders through exome sequencing. Hum. Genet. *129*, 351-370.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lee, J.H., Sung, B.H., Kim, M.S., Blattner, F.R., Yoon, B.H., Kim, J.H., and Kim, S.C. (2009). Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. Microb. Cell Fact. *8*, 2.

Lee, H.J., Kim, E., and Kim, J.S. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res. *20*, 81-89.

Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., Cheng, C.H., Dawid, I.B., Chen, Y., and Zhao, H. (2012). Efficient targeted gene disruption in *Xenopus* embryos using engineered transcription activator-like effector nucleases (TALENs). Proc. Natl. Acad. Sci. USA *109*, 17484-17489.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods *7*, 709-715.

Li, M.Z., and Elledge, S.J. (2007). Harnessing homologous recombination *in vitro* to generate recombinant DNA *via* SLIC. Nat. Methods *4*, 251-256.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res. *39*, 6315-6325.

Liu, S., Li, W., Wu, Y., Chen, C., and Lei, J. (2013). *De novo* transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. PLoS One *8*, e48156.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol. Cell *31*, 294-301.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467-477.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. Nature *470*, 198-203.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. Nat. Rev. Genet. *12*, 671-682.

Martin, C.H., Nielsen, D.R., Solomon, K.V., and Prather, K.L. (2009). Synthetic metabolism: engineering biology at the protein and pathway scales. Chem. Biol. *16*, 277-286.

Martin, J., Bruno, V.M., Fang, Z., Meng, X., Blow, M., Zhang, T., Sherlock, G., Snyder, M., and Wang, Z. (2010). Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics *11*, 663.

Matsumura, M., Takasu, N., Nagata, M., Nakamura, K., Kawai, M., and Yoshino, S. (2010). Effect of ultrafine zinc oxide (ZnO) nanoparticles on induction of oral tolerance in mice. J. Immunotoxicol. *7*, 232-237.

Matzas, M., Stahler, P.F., Kefer, N., Siebelt, N., Boisguerin, V., Leonard, J.T., Keller, A., Stahler, C.F., Haberle, P., Gharizadeh, B., et al. (2010). High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. Nat. Biotechnol. *28*, 1291-1294.

McCarroll, S.A., Huett, A., Kuballa, P., Chilewski, S.D., Landry, A., Goyette, P., Zody, M.C., Hall, J.L., Brant, S.R., Cho, J.H., et al. (2008). Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat. Genet. *40*, 1107-1112.

Medema, M.H., Breitling, R., Bovenberg, R., and Takano, E. (2011). Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. Nat. Rev. Microbiol. *9*, 131-137.

Medema, M.H., van Raaphorst, R., Takano, E., and Breitling, R. (2012). Computational tools for the synthetic design of biochemical pathways. Nat. Rev. Microbiol. *10*, 191-202.

Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D., and Wolfe, S.A. (2008). Targeted gene inactivation in zebrafish using engineered

zinc-finger nucleases. Nat. Biotechnol. *26*, 695-701.

Merlin, C., Beaver, L.E., Taylor, O.R., Wolfe, S.A., and Reppert, S.M. (2012). Efficient targeted mutagenesis in the monarch butterfly using zinc finger nucleases. Genome Res. *23*, 159-168.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*, 31-46.

Meyer, M., de Angelis, M.H., Wurst, W., and Kuhn, R. (2010). Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. Proc. Natl. Acad. Sci. USA *107*, 15022-15026.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29*, 143-148.

Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., Ozawa, T., Kodama, T., Kakeshita, H., Kageyama, Y., et al. (2008). Enhanced recombinant protein productivity by genome reduction in Bacillus subtilis. DNA Res. *15*, 73-81.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621-628.

Morton, J., Davis, M.W., Jorgensen, E.M., and Carroll, D. (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. Proc. Natl. Acad. Sci. USA *103*, 16370-16375.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501-1501.

Murakami, K., Tao, E., Ito, Y., Sugiyama, M., Kaneko, Y., Harashima, S., Sumiya, T., Nakamura, A., and Nishizawa, M. (2007). Large scale deletions in the *Saccharomyces cerevisiae* genome create strains with altered regulation of carbon metabolism. Appl. Microbiol. Biotechnol. *75*, 589-597.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Nandagopal, N., and Elowitz, M.B. (2011). Synthetic biology: integrated gene circuits. Science *333*, 1244-1248.

Ochiai, H., Fujita, K., Suzuki, K., Nishikawa, M., Shibata, T., Sakamoto, N., and Yamamoto, T. (2010). Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. Genes Cells *15*, 875-885.

Osakabe, K., Osakabe, Y., and Toki, S. (2010). Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. Proc. Natl. Acad. Sci. USA *107*, 12034-12039.

Pagani, I., Liolios, K., Jansson, J., Chen, I.M., Smirnova, T., Nosrat, B., Markowitz, V.M., and Kyrpides, N.C. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. *40*, D571-579.

Pareek, C.S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. J. Appl. Genet. *52*, 413-435.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. *10*, 669-680.

Perez-Pinera, P., Ousterout, D.G., and Gersbach, C.A. (2012). Advances in targeted genome editing. Curr. Opin. Chem. Biol. *16*, 268-277.

Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J., et al. (2009). A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. PLoS Genet. *5*, e1000569.

Pinto, A.C., Melo-Barbosa, H.P., Miyoshi, A., Silva, A., and Azevedo, V. (2011). Application of RNA-seq to reveal the transcript profile in bacteria. Genet. Mol. Res. *10*, 1707-1718.

Pleiss, J. (2006). The promise of synthetic biology. Appl. Microbiol. Biotechnol. *73*, 735-739.

Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., de Arruda, M., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. Science *312*, 1044-1046.

Pushkarev, D., Neff, N.F., and Quake, S.R. (2009). Single-molecule sequencing of an individual human genome. Nat. Biotechnol. *27*, 847-850.

Qi, L., Haurwitz, R.E., Shao, W., Doudna, J.A., and Arkin, A.P.

(2012). RNA processing enables predictable programming of gene expression. Nat. Biotechnol. *30*, 1002-1006.

Qi, X.P., Du, Z.F., Ma, J.M., Chen, X.L., Zhang, Q., Fei, J., Wei, X.M., Chen, D., Ke, H.P., Liu, X.Z., et al. (2013a). Genetic diagnosis of autosomal dominant polycystic kidney disease by targeted capture and next-generation sequencing: utility and limitations. Gene *516*, 93-100.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013b). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Qiu, Y., Cho, B.K., Park, Y.S., Lovley, D., Palsson, B.O., and Zengler, K. (2010). Structural and operational complexity of the Geobacter sulfurreducens genome. Genome Res. *20*, 1304-1311.

Quan, J., and Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One *4*, e6441.

Quan, J., Saaem, I., Tang, N., Ma, S., Negre, N., Gong, H., White, K.P., and Tian, J. (2011). Parallel on-chip gene synthesis and application to optimization of protein expression. Nat. Biotechnol. *29*, 449-452.

Ram, J.L., Karim, A.S., Sendler, E.D., and Kato, I. (2011). Strategy for microbiome analysis using 16S rRNA gene sequence analysis on the Illumina sequencing platform. Syst. Biol. Reprod. Med. *57*, 162-170.

Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. Nat. Rev. Microbiol. *10*, 607-617.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. *30*, 460-465.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Richmond, K.E., Li, M.H., Rodesch, M.J., Patel, M., Lowe, A.M., Kim, C., Chu, L.L., Venkataramaian, N., Flickinger, S.F., Kaysen, J., et al. (2004). Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. Nucleic Acids Res. *32*, 5011-5018.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods *4*, 651-657.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., et al. (2010). De novo assembly and analysis of RNA-seq data. Nat. Methods *7*, 909-912.

Rothberg, J.M., and Leamon, J.H. (2008). The development and impact of 454 sequencing. Nat. Biotechnol. *26*, 1117-1124.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat. Biotechnol. *29*, 697-698.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463-5467.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. Hum. Mol. Genet. *19*, R227-240.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*, 1086-1092.

Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J., and Church, G.M. (2000). RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. Nat. Biotechnol. *18*, 1262-1268.

Seo, J.H., Hong, J.S., Kim, D., Cho, B.K., Huang, T.W., Tsai, S.F., Palsson, B.O., and Charusanti, P. (2012). Multiple-omic data analysis of Klebsiella pneumoniae MGH 78578 reveals its transcriptional architecture and regulatory features. BMC Genomics *13*, 679.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R.,

et al. (2010). The primary transcriptome of the major human pathogen Helicobacter pylori. Nature *464*, 250-255.

Shendure, J., and Lieberman Aiden, E. (2012). The expanding scope of DNA sequencing. Nat. Biotechnol. *30*, 1084-1094.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science *309*, 1728-1732.

Shokralla, S., Spall, J.L., Gibson, J.F., and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. *21*, 1794-1805.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X., et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature *459*, 437-441.

Sleight, S.C., Bartley, B.A., Lieviant, J.A., and Sauro, H.M. (2010). In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res. *38*, 2624-2636.

Smith, M.G., Gianoulis, T.A., Pukatzki, S., Mekalanos, J.J., Ornston, L.N., Gerstein, M., and Snyder, M. (2007). New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev. *21*, 601-614.

Soon, W.W., Hariharan, M., and Snyder, M.P. (2013). High-through-put sequencing for biology and medicine. Mol. Syst. Biol. *9*, 640.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol. *6*, 181-186.

Sung, P., and Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat. Rev. Mol. Cell Biol. *7*, 739-750.

Surget-Groba, Y., and Montoya-Burgos, J.I. (2010). Optimization of de novo transcriptome assembly from next-generation sequencing data. Genome Res. *20*, 1432-1440.

Takasu, Y., Kobayashi, I., Beumer, K., Uchino, K., Sezutsu, H., Sajwan, S., Carroll, D., Tamura, T., and Zurovec, M. (2010). Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. Insect Biochem. Mol. Biol. *40*, 759-765.

Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X., Zhang, L., et al. (2011). Knockout rats generated by embryo microinjection of TALENs. Nat. Biotechnol. *29*, 695-696.

Thiele, I., Swainston, N., Fleming, R.M., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. Nat. Biotechnol. [Epub ahead of print]

Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004). Accurate multiplex gene synthesis from programmable DNA microchips. Nature *432*, 1050-1054.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., and Voytas, D.F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature *459*, 442-445.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646-651.

Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R.M., and Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat. Methods. *5*, 829-834.

van der Ploeg, J.R. (2009). Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology *155*, 1966-1976.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J.,

Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304-1351.

Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. Nature *388*, 97-100.

Wang, J., Friedman, G., Doyon, Y., Wang, N.S., Li, C.J., Miller, J.C., Hua, K.L., Yan, J.J., Babiarz, J.E., Gregory, P.D., et al. (2012a). Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. Genome Res. *22*, 1316-1326.

Wang, L., Tsutsumi, S., Kawaguchi, T., Nagasaki, K., Tatsuno, K., Yamamoto, S., Sang, F., Sonoda, K., Sugawara, M., Saiura, A., et al. (2012b). Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. Genome Res. *22*, 208-219.

Wang, Y.H., Wei, K.Y., and Smolke, C.D. (2013a). Synthetic biology: advancing the design of diverse genetic systems. Ann. Rev. Chem. Biomol. Eng. [Epub ahead of print]

Wang, R.Y., Shi, Z.Y., Guo, Y.Y., Chen, J.C., and Chen, G.Q. (2013b). DNA fragments assembly based on nicking enzyme system. PLoS One *8*, e57943.

Westers, H., Dorenbos, R., van Dijl, J.M., Kabel, J., Flanagan, T., Devine, K.M., Jude, F., Seror, S.J., Beekman, A.C., Darmon, E., et al. (2003). Genome engineering reveals large dispensable regions in Bacillus subtilis. Mol. Biol. Evolution *20*, 2076-2090.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. Nature *452*, 872-876.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331-338.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., and Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature *453*, 1239-1243.

Wold, B., and Myers, R.M. (2008). Sequence census methods for functional genomics. Nat. Methods *5*, 19-21.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., et al. (2011). Targeted genome editing across species using ZFNs and TALENs. Science *333*, 307.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat. Struct. Mol. Biol. *16*, 130-137.

Yu, B.J., Sung, B.H., Koob, M.D., Lee, C.H., Lee, J.H., Lee, W.S., Kim, M.S., and Kim, S.C. (2002). Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat. Biotechnol. *20*, 1018-1023.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T., et al. (2010). High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. Proc. Natl. Acad. Sci. USA *107*, 12028-12033.

Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011). The impact of next-generation sequencing on genomics. J. Genet. Genomics *38*, 95-109.

Zhang, Y., Werling, U., and Edelmann, W. (2012). SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. *40*, e55.

Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q., et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell *5*, 97-110.