Supplementary Note

Optimization of OLS Pool 2 assemblies.

We found that three factors were critical for the robust assembly of OLS Pool 2 subpools into full-length products (data not shown). First, using a pre-assembly step of 15-20 thermal cycles performed in the absence of construction primers followed by 20-30 cycles of assembly PCR with the construction primer reduced formation of incorrect products. Second, using low annealing temperatures (50-55 °C) during the pre-assembly and more stringent annealing temperatures during the assembly PCR (60-72 °C) also helped reduce the formation of secondary bands. Third, using an increased DNA concentration (5-20 ng/μL) in the pre-assembly step (2-3 orders of magnitude greater than for OLS Pool 1) was the most important factor in improving the reliability between different synthesis products.

Differences in quantity of OLS material utilized between OLS Pools 1 and 2.

While the 13,000 features in OLS Pool 1 can be used to construct >700 genes, each subpool amplification used 1/500th of the total chip-eluted DNA. While it maybe possible to run this process with 1/1000th the total material, we were concerned that the use of larger OLS Pools would be difficult (e.g., a 55,000 feature OLS pool would require 1/3,000th of the total material). The longer 200mers of OLS Pool 2 allowed for a first plate amplification before the assembly amplification reducing the OLS Pool 2 usage by two orders of magnitude, which facilitates process scaling to larger OLS Pools.

Summary of previous chip-based synthesis efforts

Previous efforts have demonstrated the ability to synthesize genes from DNA microchips, though in limited pool complexities. Tian et al. described the assembly of 14.6 kilo-basepairs of novel DNA from 292 oligonucleotides synthesized on an Atactic/Xeotron chip[1]. The process involved using 584 shorter oligonucleotides synthesized on the same chip for hybridization-based error correction. Even with the 584 shorter oligonucleotides, the total background oligonucleotide size was less than 50 kilobases. The resulting error rates were ~1/160 bp before error correction and ~1/1400 bp after. Using similar chips, Zhou et al. constructed ~12 genes with an error rate as low as 1/625 bp[2]. Richardson et al. showed the assembly of an 180 bp construct from 8 oligonucleotides synthesized on a microarray using maskless photolithographic deprotection (Nimblegen)[3]. The error rates were not determined, though a follow-up construction of a 742 bp GFP sequence

Nature Biotechnology: doi:10.1038/nbt.1716

using the same process showed an error rate of 1/20 bp – 1/70 bp[4]. The background oligonucleotide complexity of this last study was not published. The oligonucleotide pool complexity for OLS Pool 2 was a total of ~2.5 megabases of DNA.

Nature Biotechnology: doi:10.1038/nbt.1716

Supplementary Methods

OLS Pool 1

Overall Design

The first OLS library (OLS Pools 1) was designed primarily for experiments not described in this paper, and it consisted of 12 separately amplifiable assembly subpools. Of the 13,000 oligonucleotides (oligos) that were made in OLS Pool 1, there were two subpools, GFP43 and GFP35, that were designed to each synthesize the mut3 variant of GFP (GFPmut3b)[5]. GFP43 consisted of 18 oligos while GFP35 had 22. The individual subpools assembled into 779 bp constructs, of which 719 bp could be cloned and verified downstream after restriction digest. Two other subpools were used as amplification controls (Control 1 and 2) and contained 10 and 5 130mers, respectively. The remaining 12,945 OLS Pool 1 oligos consisted of 130mers having homology to the *E. coli* genome that were split into 8 separate amplification subpools (sequences available upon request). The OLS array was synthesized, processed from the chip, and delivered as a ~1-10 pmol lyophilized pool of oligos by Agilent Technologies (Carlsbad, CA).

Design of GFPmut3 Assembly Subpools

We designed forward and reverse GFPmut3 assembly oligos to have complete overlap, as well as a bridging oligonucleotide to allow for tests with both circular ligation assembly (data not shown) and PCR assembly protocols[10]. The overlap lengths were 43 bp and 35 bp for GFP43 and GFP35, respectively. We developed an algorithm that automatically splits the constructed sequences into adjacent annealing segments of similar melting temperatures that was loosely based on the Gene2Oligo design method[11]. Briefly, the algorithm first adds random DNA sequence on the ends of the constructed gene to allow for leeway on the first and last annealing segment. Next, the algorithm enumerates all possible overlap regions for the gene to be constructed that fall within a certain length range and sorts them into bins based on their start position. The mean melting temperature is calculated for all overlap regions, and regions that do not fall within a defined temperature deviation are removed. Bins are sorted in order based on minimal deviation from the mean melting temperature. The program then recursively attempts to construct the gene from left to right by picking the first region from the top of the list. If a particular position has no annealing regions (no regions match the melting temperature), the program backtracks and picks the next valid annealing region and tries again. Once a valid set of annealing regions is designed, the algorithm designs oligos that span two adjacent annealing regions alternating between the sense and antisense strands. Finally, a bridging oligo that spans the first and last segment is designed. The requirement of a bridging oligo necessitates that

Nature Biotechnology: doi:10.1038/nbt.1716

an even number of annealing regions are designed and the algorithm takes this into account.

The GFP43 subpool was designed using a seed overlap region size of 43, size variability of ±2, and a temperature variability of 4.5 °C. The resultant designs had 18 oligos with a mean melting temperature of 72.5 °C with a 1.8 °C average deviation. The GFP35 subpool was designed using a seed overlap region size of 35, size variability of ±4, and temperature variability of 3 °C. The resultant designs had 22 oligos with a mean melting temperature of 69.6 °C with a 1.6 °C average deviation. Finally, we designed a pool of oligos, GFP20, that were made using column-based synthesis and which could construct GFPmut3. The GFP20 design used a seed overlap region size of 20, size variability of 3, and a temperature variability of 5 °C. The resultant designs had 40 oligonucleotides with a mean melting temperature of 56.3 °C with a 1.0 °C average deviation.

Design of Subpool Assembly-Specific Primers

There were a total of 12 assembly subpools designed for OLS Pool 1. We selected orthogonal primers from a set of 240,000 previously designed orthogonal 25mer barcodes designed for yeast genomic hybridization studies[8]. Briefly, each barcode was searched for reverse primers for 20mers that end in 'GATC'. Forward primers were selected from barcode primers that end in 'T'. Both forward and reverse primer sets were screened for melting temperatures between 62 °C and 64 °C calculated using the nearest neighbor method[9,10]. Primers were then screened by BLAT for hits (tilesize=6, stepsize=1,minMatch=1) against one another, as well as against the E. coli genome[11]. Primers with greater than 1 self-hit, or 3 *E. coli* genome hits were removed. Secondary structures were then calculated using UNAFold, and any primers containing folding energies less than 0 kcal/mol were removed[12]. Primers pairs were then screened using MultiPLX to obtain a group of orthogonal primers, from which 12 primers were chosen to be assembly-specific primers[13]. All scripts (available upon request) were written in Python and used several BioPython utilities[14].

Assembly Subpool Amplification

Lyophilized DNA recovered from OLS Pool 1 (~1 pmol total DNA) was resuspended in 500 µL TE Buffer. Each of the four assembly subpools (GFP43, GFP35, Control 1, and Control 2) were amplified in 50 µL reactions using the *KAPAprep* protocol (all italicized PCR protocols are named and described in the PCR protocol Table at the end of this supplement) with the appropriate assembly-specific primers and 1uL of the reconstituted OLS Pool 1. These PCR reactions were monitored by real-time PCR and were stopped before reaching plateau fluorescence levels to prevent over-amplification

Nature Biotechnology: doi:10.1038/nbt.1716

(between 35-45 cycles). Two replicates were pooled and purified using QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA). The resultant subpools were size verified and quantified on gels to give between 20 and 35 ng/μL of DNA in 30 μL total. 20 μL of each subpool was re-amplified in 20 mL total volume spread split into two 96-well plates using the *TaqPrep* protocol with chemically modified assembly-specific primers (see Supplementary Figure 11 for details). Samples were spun down in Amicon Ultra-15 mL Centrifugal Filter with Ultracel-10 membrane at 4,000 g in a swinging bucket rotor, washed in 13mL TE Buffer, and recovered into 350 μL total volume. 40μL of 1AU/mL QIAGEN Protease was added to each sample, and shaken at 800 rpm in a Thermomixer R (Eppendorf AG, Hamburg Germany) at 37 °C for 40 min, and then 20 min at 70 °C to heat inactive. 70 μL of RapidClean Protein Removal Resin (Advantsa, Menlo Park CA) was added, mixed for 15 s, and spun down at 24,000 g in an Eppendorf Centrifuge 5424 for 5 min, and the supernatant was removed. The supernatant was rewashed in water in an Amicon Ultra-0.5mL Centrifugal Filter with Ultracel-10 membrane and volume adjusted to 450 μL.

Assembly Subpool Processing
Purified samples from above were treated with Lambda Exonuclease (Enzymatics) to make them single stranded. 445 μL of the filtrate, 150 μL 10x Lambda Exonuclease Buffer, 805 μL water, and 100 μL Lambda Exonuclease was incubated at 37 °C for 40 min and 20 °C for 20 min and shaken at 800 rpm in a Thermomixer R. Samples were spun down in Amicon Ultra-0.5mL Centrifugal Filter with Ultracel-3 membrane and washed with water and recovered in 350μL water. 300μL of each sample were then processed with 1250U of DpnII (New England Biolabs, Ipswitch, MA), 125U USER Enzyme (New England Biolabs), and 3 nanomoles of the guide oligo (the reverse subpool amplification primer without a 5' phosphate) in 2.5 mL of 1x DpnII buffer, and incubated at 800 rpm at 37 °C. Samples were then filtered in an Amicon Ultra-15mL 3kDA filter, washed first with 2 mL TE, and then with 4 mL water. The ssDNA product was recovered in 130 μL for control subpools 1 and 2, and 50 μL for GFP43 and GFP35 assembly subpools.

First OLS Pool 1 Assemblies

Assembly
GFPmut3b was assembled from column-synthesized oligos (IDT, Coralville, IA) by amplifying 1 μL of a pool of 19 reverse oligos (1.05 μM each) and 20 forward oligos (1 μM each) in a 20 μL reaction using the *Phu1* protocol with the forward and reverse construction primers (GFPfwd and GFPrev, IDT). The reaction was heated to 98 °C for 30 s, followed by 30 cycles of 98 °C for

Nature Biotechnology: doi:10.1038/nbt.1716

5 s, 51 °C for 10 s, and 72 ° for 30 s. This was followed by a final extension of 72 °C for 10 min.

The concentrations of the assembly subpools were determined using a Nanodrop 2000c spectrophotometer (Thermo Scientific, Wilmington, DE), as were all measurements of DNA concentration described in the methods henceforth. GFP43 and GFP35 assembly subpools were assembled into GFPmut3 by amplifying 330 pg of GFP43 or 40 pg of GFP35 in a 20 μL reaction using the *Phu1* protocol with the forward and reverse construction primers (GFPfwd and GFPrev). The full-length products from both assemblies were isolated by running 18 μL of the assembly PCR on four lanes of a 2% EX E-Gel (Invitrogen, Carlsbad, CA) and extracting the DNA using a QIAquick Gel Extraction Kit (QIAGEN). This yielded 4 ng and 6 ng of GFPmut3 built from subpools GFP43 and GFP35, respectively – both in 50 μL EB buffer (10 mM Tris-Cl, pH 8.5). 1 μL of the gel-isolated DNA was amplified in 20 μL reactions using the *Phu1* protocol. Each gel-isolated assembly was amplified in 24 different PCR reactions. The amplification products were cleaned up using a QIAquick PCR Purification Kit.

Cloning
For screening all fluorescent proteins, we used the expression plasmid pZE21[15]. NEB 10-beta (New England Biolabs) *E. coli* cells transformed with the plasmid were streaked out on LB agar plates containing 50 μg/mL kanamycin. A single colony was then grown for 17 hr in 2 mL LB with 50 μg/mL kanamycin and thereafter kept at 4 °C for <60 h. This culture was back-diluted by adding 100 μL to 100 mL of fresh LB/kanamycin medium and grown for 17 h at 37 °C and stored at 4 °C for 3 h. The plasmid was isolated using QIAprep Spin Miniprep Kit (QIAGEN)

GFPmut3b was amplified from 9-10 ng of pZE21G[16] in 50 μL reactions using the *Phu2* protocol with the primers GFPfwd2 and GFPrev2. The products were cleaned up using a QIAquick PCR Purification Kit. To generate the stock of control GFPmut3 used in all subsequent fluorescent protein cloning experiments, 10-20 ng of the amplified product was re-amplified in 50 μL reactions using the *Phu2* protocol (except that dNTPs from Kapa Biosystems were used), again using primers GFPfwd2 and GFPrev2. The products were cleaned up using a QIAquck PCR Purification Kit.

4.9 μg of GFP43 assembly, 5.8 μg of GFP35 assembly, 4.2 μg of GFP20 assembly, 2.7 μg of the GFP control, and 2.7 μg of pZE21 were digested in separate 50 μL reactions that consisted of 1x NEBuffer 2 (500 mM NaCl, 100 mM Tris-HCl, 100 mM $MgCl_2$, 10 mM dithiothreitol, pH 7.9; New England Biolabs), 100 ng/μL bovine serum albumin (New England Biolabs), 0.4

Nature Biotechnology: doi:10.1038/nbt.1716

units/μL of HindIII (New England Biolabs), and 0.54 units/μL KpnI (New England Biolabs). The assemblies were digested at 37 °C for 3 h while shaking at 800 rpm in a Thermomixer R. After GFP control and pZE21 were digested for 2.5 hrs at 37 °C, 1 μL of 20 units/μL DpnI (New England Biolabs) was added to the GFP control digests and 1 μL of 5 units/μL Antarctic phosphatase (New England Biolabs) and 5.6 μL 10x Antarctic phosphatase buffer (New England Biolabs) were added to the pZE21 digests. The GFP control and plasmid were kept at 37 °C for 30 min while shaking at 800 rpm in a Thermomixer R. The enzymes in all reactions were heat inactivated at 65 °C for 20 min while shaking at 800 rpm in a Thermomixer R. The products were cleaned up using a QIAquick PCR Purification Kit.

HindIII/KpnI digested assemblies from GFP43, GFP35, or GFP20 were cloned as follows. 180 ng of one of the inserts and 40 ng of HindIII/KpnI digested pZE21 were diluted in 8.5 μL water. 1 μL of 10x T4 ligase buffer (New England Biolabs) was added, and the reaction was heated to 37 °C for 5 min. The reaction was brought down to room temperature, and 0.5 μL of 400 units/μL of T4 DNA ligase (New England Biolabs) was rapidly added. The ligation was then allowed to proceed for 10 min at 25 °C. The enzyme was heat-inactivated for 15-25 min at 65 °C. All thermal steps were conducted with shaking at 800 rpm in a Thermomixer R. A 25 nm mixed cellulose ester membrane (Milipore, Billerica, MA) was used to dialyze the ligation product against a 1,000-fold greater volume of water for 5-15 min. 2 μL of the dialyzed ligation product was added to 50 μL freshly thawed NEB 10-beta electrocompetent *E. coli* cells (New England Biolabs*)*, and the mixture was briefly incubated on ice. Electroporation was performed with one pulse of 1.8 kV using Gene Pulser cuvettes with a 0.1 cm electrode gap (Bio-Rad, Hercules, CA) in a MicroPulser (Biorad). The cells were suspend in 1 mL LB medium and incubated at 37 °C for 70 min. A fraction of each culture was then plated onto 50 μg/mL kanamycin LB agar plates and grown overnight at 37 °C. The 1 mL non-selective culture was stored at 4 °C for 23 hr, after which 1 μL was inoculated into 1 mL of 50 μg/mL LB that was subsequently grown overnight at 37 °C.

Flow Cytometry

For each cloning reaction, 10 μL of the ovenight culture in selective medium was added to 1 mL 50 μg/μL kanamycin and grown at 37 °C for 1-2 hr. The fluorescent cell fraction was then quantified using a BD LSRFortessa flow cytometer (BD Biosciences, San Jose, CA) using a 488 nm blue laser and a FITC detector (530 nm filter with 30 nm bandpass).

Sequencing

Nature Biotechnology: doi:10.1038/nbt.1716

Colonies were randomly picked from selective agar cultures corresponding to each ligation reaction. Each colony was inoculated into 200 μL of 50 μg/mL LB and grown overnight at 32 °C. Each 200 μL overnight culture was split into two 100 μL aliquots, and 100 μL 30% glycerol in water was added to each aliquot. The stocks were immediately placed into -80 °C storage. Dideoxy sequencing of one of the two 200 μL glycerol stocks was performed by Beckman Coulter Genomics (Danvers, MA) using the following primers: forward – 5'ataaaaataggcgtatcacgaggc; reverse – 5'cggcggatttgtcctactcag. The second glycerol stock was kept to make possible the recovery of sequenced clones.

Second OLS Pool 1 Assemblies

Assembly
170 pg of the GFP43 and 190 pg of the GFP35 assembly subpools were assembled into GFPmut3 in separate 20 μL reactions using the *Phu1* protocol with the construction primers (GFPfwd and GFPrev). The full length products were isolated from a 2% agarose gel using a QIAquick Gel Extraction Kit, with the product of 23 GFP43 assembly reactions concentrated into 50 μL EB buffer, and 70 GFP35 assembly reactions concentrated into 135 μL EB buffer. 10 μL of the assembly products were then digested in 50 μL KpnI/HindIII reactions identical to the one described during the cloning of the first set OLS Pool 1 assemblies (except for the lack of the 65 °C heat inactivation step). The digested products were cleaned up using a MinElute PCR Purification Kit (QIAGEN).

Cloning
Using a 2% EX E-Gel and a quantitative DNA ladder, the concentrations of GFPmut3 assemblies from GFP43 and GFP35 were determined to be 14 ng/μL and 35 ng/μL, respectively. The PCR-amplified KpnI/HindIII-digested 40 ng/μL GFPmut3 stock prepared during the first assembly experiment was used as a positive control, and the 180 ng/uL stock of KpnI/HindIII-digested pZE21 prepared during the same experiment was used as the cloning vector. Electrocompetent *E. coli* cells were prepared by concentrating a 2 L culture of NEB 5-alpha cells (New England Biolabs) into 50 mL of water.

14ng of GFP43 and 35ng of GFP35 were each added to 180ng of vector and were ligated in a 10 μL T4 ligase reaction the products of which were electroporated into NEB 5-alpha cells following the protocol described in the cloning of the first OLS Chip 1 constructs. After an outgrowth of 37 °C for 70 min, 100 μL of the culture was diluted into 900 μL of LB with 50 μL/mL kanamycin, and another fraction was plated onto 50 μg/mL kanamycin LB

Nature Biotechnology: doi:10.1038/nbt.1716

agar plates. Both the plated cells and the cells in liquid culture were grown overnight at 37 °C.

Flow Cytometry
20 μL of each overnight culture of the non-error corrected constructs was diluted into 2 mL 50 μg/μL kanamycin LB and grown at 37°C for 2 hrs. The fluorescent cell fraction was then quantified using a BD LSRFortessa.

Sequencing
Random clones were grown overnight in LB, made into glycerol stocks, and sequenced by Beckman Coulter Genomics following the protocol described in the sequencing of the first OLS Chip 1 constructs.

Error Correction
HindIII/DpnI-digested assemblies (840 pg of GFP43 and 380 pg of GFP35) were amplified in separate 20 μL reactions following the *Phu3* protocol and using the primers GFPfwd3 and GFPrev3. Each assembly was amplified in four 20 μL reactions, which were subsequently pooled and cleaned up in a single QIAquick PCR Purification Kit column.

Error correction using ErrASE (Novici Biotech, Vacaville, CA) was performed using a slight variation of the manufacturer's protocol. In brief, either 2.8-2.9 μg of GFP protein assembly were added to separate 50 μL reactions consisting of 0.9x Phusion HF buffer with 180 μM dNTPs (Enzymatics). Each reaction was heated to 98 °C for 1 min, cooled to 0 °C for 5 min, kept at 37 °C for 5 min, and subsequently stored and handled at 4 °C. 10 μL of the reaction was then added to each of first five of the six decreasingly stringent ErrASE reactions, and the mix was incubated at 25 °C for 1 h while shaking at 800 rpm in a Thermomix R. 2 μL of the ErrASE reactions were then re-amplified in 50 μL reactions using the *Phu3* protocol with the primers GFPfwd3 and GFPrev3.

Post-ErrASE Cloning, Flow Cytometry, and Sequencing
The highest stringency ErrASE reaction that resulted in a PCR product (#2 for both assemblies) was cleaned up using a QIAquick PCR Purification Kit. 260 ng of GFP43 and 960 ng of GFP35 were digested in 40 μL reactions with 4 μL NEBuffer 2, 0.4 μL bovine serum albumin, 0.5 μL HindIII (20 units/μL), 1.4 μL KpnI (10 units/μL), and water. The error-corrected constructs were digested at 37 °C for 2 h while shaking at 800 rpm in a Thermomixer R. Although electrophoresis on an agarose gel detected only the single, correct band, we nonetheless gel-isolated the constructs using a QIAquick Gel Extraction Kit in order to remove any undetected misassemblies.

Nature Biotechnology: doi:10.1038/nbt.1716

20 ng of pZE21 and either 35 ng of gel-isolated GFP43, 65 ng of gel-isolage GFP35, or 70 ng of control GFP (same prep as was used during the previous ligation experiments) were diluted in 8.5 μL water. The DNA was then ligated in a 10 μL T4 ligase reaction the products of which were electroporated into NEB 5-alpha cells following the protocol described in the cloning of the first OLS Chip 1 constructs. After an outgrowth of 37 °C for 65 min, 400 μL of the culture was diluted into 2 mL of LB with 50 μL/mL kanamycin, and another fraction was plated onto 50 μg/mL kanamycin LB agar plates. Both the plated cells and the cells in liquid culture were grown overnight at 37 °C.

For each overnight culture, 5 μL was diluted into 500 μL 50 μg/μL kanamycin LB and grown at 37 °C for 1.5 h. The fluorescent cell fraction was then quantified using the BD LSRFortessa flow cytometer. The fluorescent fraction of each overnight culture was measured across 7-8 technical replicates. The data from one replicate per culture was removed from the analysis due to obvious fluidics-mediated sample carryover between the last wells and the first wells of the different experiments conditions.

Random clones were grown overnight in LB, made into glycerol stocks, and sequenced by Beckman Coulter Genomics following the protocol described in the sequencing of the first OLS Chip 1 constructs (except that the overnight culture was performed at 37 °C).

OLS Pool 2

Overall design
The pool of oligos from the second OLS chip (OLS Pool 2) was designed specifically for gene synthesis applications. In total, the chip encoded 12,998 oligonucleotides encoding 2,456,706 nucleotides of synthetic DNA. OLS Pool 2 was split into 11 plate subpools, which were further divided into a total of 836 assembly subpools. The 836 potential assemblies encoded 869,125 bp of DNA after all primer processing steps.

Redesign of Orthogonal Primers
We initially began by scaling up the primer design method for OLS Pool 1 to allow for the design of 3,000 orthogonal primer pairs. Again, we began with the same set of 240,000 orthogonal barcodes as in OLS Pool 1. In order to facilitate current and possible future downstream cloning and processing steps, we first removed primers containing restriction enzyme recognitions sites to the following enzymes: AatII, BsaI, BsmBI, SapI, BsrDI, BtsI, EarI, BspQI, BbsI, BspMI, BfuAI, NmeAIII, BamHI, NotI, EcoRI, KpnI, HindIII, XbaI, SpeI, PstI, PacI, and SbfI. We then removed all primers with melting temperature below 60 °C and above 64 °C to facilitate melting temperature

Nature Biotechnology: doi:10.1038/nbt.1716

matching of forward and reverse primers. Finally, we implemented an algorithm that screens primers for primer dimer formation that follows the AutoDimer program[17], though giving double weight to the terminal 10 bases on the 3' end. All primers with a score > 3 were removed. After these screens, 155,608 primers remained. A BLAST library was constructed of all synthesized genes on the chip (except the fluorescent proteins), and each oligo was screened against the library using BLAT (tileSize=6, stepSize=1, minMatch=2, maxGap=4) and any primers with hits were removed leaving 70,498 primers. A second BLAST library was constructed from the remaining primers, and a network elimination algorithm as described in the orthogonal barcode paper was applied (tileSize=6, stepSize=1, -minMatch=1, maxGap=4)[8]. This resulted in 8275 remaining primers, which were screened for formation of secondary structure ($\Delta G>-2$). Finally, the 7738 remaining primers were aligned using clustalw2 (default options for DNA(slow)), clustered, and a phylogenetic tree was generated. This tree was traversed to find 200 nodes that were distant from one another and contained at least 30 primers each. Then, one primer from each batch was chosen. Primers were sorted on melting temperature, and then paired provided that they pass a primer dimer test (filtered dimers with a score >4). The final output is a set of 3,000 pairs of orthogonal primers, grouped in sets of 100. The first set was reserved as plate-specific primers (skpp1-100), the second set was reserved for contruction primers (skpp101-200), and each remaining set was used in order for assembly-specific primers.

Construct Designs

We wrote automated algorithms to split constructs into oligonucleotide segments with partial overlaps to allow for stringent PCR assembly. Given a desired overlap size, allowable leeway on the size and position of the overlaps, and a melting temperature range, and TypeIIs restriction enzyme site, the program automates the process of turning full-length gene constructs into oligonucleotides to be synthesized on the OLS platform. Briefly, the algorithm starts by padding the sequence with the proper construction primers. Then, the construct is evenly divided among the number of necessary oligonucleotides to construct the whole sequence, automatically determining the starting overlap positions. These overlap positions are screened for melting temperature falling within the defined length range, secondary structure formation (($\Delta G>-3$), and self dimer formation (score>3) (see orthogonal primer design section). If these conditions are not met, the overlap lengths and positions are progressively varied and rechecked according to the predefined boundaries set at the beginning of the run. Once an overlap set is found that satisfy all the conditions, the final oligonucleotides are defined, and then flanked with the proper TypeIIs resctriction sites followed by the assembly-specific and plate-specific primer sequences. All sequences are rechecked for proper

Nature Biotechnology: doi:10.1038/nbt.1716

restriction enzyme cutting to make sure additional restriction sites were not added by adding primer sequences (in which case, the program pads with arbitrary sequence to remove the restriction site).

64 assemblies were designed that encoded three codon-optimized fluorescent proteins, mTFP1[18], mCitrine[19], and mApple[20]. Codon-optimization was done using a custom algorithm that randomly assigned codons weighted to their natural frequencies in the *E. coli* genome as defined by the Kazusa Codon Usage Database (http://www.kazusa.or.jp/codon/). Each protein (mApple was synthesized twice for each of these conditions) was fed through the algorithm varying overlap length (15,18,22,25 bp average overlaps) and fixing TypeIIs cutters (BtsI and BspQI), or varying TypeIIs restriction enzyme sites (BtsI, BspQI, BsrDI, EarI, BsaI, BsmBI, SapI, BbsI) and fixing average overlap lengths (20). The allowable melting temperature ranges were: 15bp overlap - 50-55 °C; 18bp overlap – 53-58 °C; 20bp overlap – 58-62 °C; 22bp overlap – 58-65 °C; 25bp overlap – 65-72 °C. In addition, the overlap length leeway was set to ±3, and position leeway to ±5. These 64 assemblies were designed to be amplified together using a single plate-specific amplification, and then individually using assembly-specific primers. We only show the assembly of one of the conditions in this work, which is from the BtsI with 20bp overlap conditions.

The 42 antibody assemblies were designed as described in the main text (V region sequences were obtained from the IMGT database[21]). Amino acid sequences for the antibodies were codon optimized for human expression using the same algorithm and settings as the fluorescent protein designs in the 20bp overlap, BtsI restriction enzyme condition. The segments of the 42 antibodies were flanked by different plate-specific pool primers than the fluorescent proteins, and individually addressable using assembly-specific primers.

Fluorescent Proteins from OLS Pool 2

Amplification of Plate and Assembly Subpools
As with the OLS Pool 1, oligos were synthesized, processed from the chip, lyophilized, and then reconstituted in 500 μL TE buffer. Plate subpools were made by amplifying 1 μL of OLS Pool 2 oligos in 50 μL reactions with the *Phu4* PCR protocol using the forward and reverse plate-specific primers (skpp1F and skpp1R). Fluorescent protein assembly subpools pools were amplified from the plate pool by including 20 nL of the plate subpool in 100 μL reactions that used the *Phu4* protocol (except that the number of cycles was increased to 30) with the correct forward and reverse assembly-specific primers (skpp201F-skpp204F and skpp201R-skpp204R). The products were

Nature Biotechnology: doi:10.1038/nbt.1716

cleaned up using a QIAquick PCR Purification Kit, with the elution step conducted using 0.25x EB buffer diluted in water. The resulting DNA concentration of the assemblies was ~90 ng/uL.

Assembly
2 μL of each fluorescent protein assembly subpool were "pre-assembled" in 20 μL reactions following the *KODpre* protocol. 5 μL of each pre-assembly reaction was then assembled in 50 μL reactions following the *KOD1* protocol and using the appropriate forward and reverse construction primers (skpp101F-skpp142F and skpp101R-skpp142R). The products were cleaned up using a MinElute PCR Purification Kit.

Cloning
180 ng of mTFP1assembly, 1.6 μg of mCitrine assembly, or 190 ng of mApple assembly were digested with HindIII and KpnI in 50 μL reactions identical to the one described for the cloning of the OLS Pool 1 constructs (except that the length of digest was 2 h rather than 3 h). The digested products were cleaned up using a MinElute PCR Purification Kit.The PCR-amplified KpnI/HindIII-digested 40 ng/μL GFPmut3 stock prepared during the first OLS Pool 1 assembly experiment was used as a positive control, and the 180 ng/uL stock of KpnI/HindIII-digested pZE21 prepared during the same earlier experiment was used as the cloning vector. Electrocompetent *E. coli* cells were prepared by concentrating a 2 L culture of NEB 5-alpha cells into 50 mL of water.

40 ng of pZE21 and either 60 ng of mTFP-BtsI-20 assembly, 90 ng of mCitrine-BtsI-20 assembly, 30 ng of mApple-BtsI-20, or 180 ng of control GFP were diluted in 8.5 μL water. The DNA was then ligated in a 10 μL T4 ligase reaction the products of which were electroporated into NEB 5-alpha cells following the protocol described in the cloning of the first OLS Chip 1 constructs. After an outgrowth of 37 °C for 70 min, 100 μL of the culture was diluted into 900 μL of LB with 50 μL/mL kanamycin, and another fraction was plated onto 50 μg/mL kanamycin LB agar plates. Both the plated cells and the cells in liquid culture were grown overnight at 37 °C.

Flow Cytometry
For each overnight culture, 20 μL was diluted into 2 mL 50 μg/μL kanamycin LB and grown at 37°C for 2-3 hrs. The fluorescent cell fraction was then quantified using a BD LSRFortessa flow cytometer.

Optimizing ErrASE

Error Correction

Nature Biotechnology: doi:10.1038/nbt.1716

Error correction using ErrASE was performed using the manufacturer's instructions. In brief, 2.4 μg of each fluorescent protein assembly (described above) were added to separate 60 μL reactions consisting of KOD polymerase buffer with 200 μM NTPs (EMD Chemicals) and 1.46 μM $MgSO_4$. Each reaction was heated to 98 °C for 1 min, cooled to 0 °C for 5 min, kept at 37 °C for 5 min, and subsequently stored and handled at 4 °C. 10 μL of the reaction was then added to each of the six ErrASE reactions of decreasing stringency, and the mix was incubated at 25 °C for 1-2 h. The ErrASE reactions were then re-amplified by adding 2 μL to a 50 μL amplification reaction identical to KOD PCR used to assemble the fluorescent proteins.

Cloning
Following error correction the amplifications that produced a band the size of a full-length assembly were cleaned up using a QIAquick PCR Purification Kit, with the DNA eluted into 30 μL of EB buffer. The error-corrected products were then digested with HindIII and KpnI in 50 μL reactions identical to the one described for the cloning of the OLS Pool 1 constructs. The digest was done at 37 °C for 3 h while shaking at 800 rpm in a Thermomixer R. The digested products were cleaned up using a MinElute PCR Purification Kit. The PCR-amplified KpnI/HindIII-digested 40 ng/μL GFPmut3 stock prepared during the first OLS1 assembly experiment was used as a positive control, and the 180 ng/uL stock of KpnI/HindIII-digested pZE21 prepared during the same earlier experiment was used as the cloning vector. Electrocompetent *E. coli* cells were prepared by concentrating a 2 L culture of NEB 5-alpha cells into 50 mL of water.

40 ng of pZE21 and 100-180 ng/μL of the inserts were ligated in a 10 μL T4 ligase reaction the products of which were electroporated into NEB 5-alpha cells following the protocol described in the cloning of the first OLS Chip 1 constructs. After electroporation the cells were outgrown in 1 mL of non-selective LB for 37 °C for 70 min, of which 100 μL was diluted into 900 μL of 50 μg/mL kanamycin LB and grown overnight at 37 °C.

Flow Cytometry
For each overnight culture, 20 μL was diluted into 2 mL 50 μg/μL kanamycin LB and grown at 37°C for 2-3 hrs. The fluorescent cell fraction was then quantified using a BD LSRFortessa flow cytometer using both a 488 nm blue laser with a FITC detector (530 nm filter with 30 nm bandpass) and a 561 nm yellow laser with a Texas Red detector (610 nm filter with a 20 nm bandpass).

Antibodies from the second OLS chip – first set of assemblies

Nature Biotechnology: doi:10.1038/nbt.1716

Amplification and processing of antibody assembly pools
Plate-specific assembly pools were amplified from the full set of 12,998 OLS2 oligos in 50 μL *Phu4* reactions with 1 μL OLS and using the plate-specific amplification primers skpp2F and skpp2R.

To make antibody assembly subpools, 20 ng of the plate subpool was amplified in 100 μL reactions following the *Phu5* protocol and using the appropriate forward and reverse amplification primers (skpp301F-skpp342F and skpp301R-skpp342R). The reaction was cleaned up using a QIAquick PCR Purification Kit, with each 100 μL reaction concentrated into 30 μL EB buffer.

30 μL of the amplified antibody assembly subpools were digested with BtsI in 40 μL reactions with 1x NEBuffer 4 (50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9; New England Biolabs), 125 ng/μL bovine serum albumin (New England Biolabs), and 0.5 units/μL BtsI (New England Biolabs). The reaction was cleaned up using a MinElute PCR Purification Kit.

Assembly optimization
125 ng of each antibody assembly subpool were pre-assembled in separate 20 μL reactions following the *KODpre* protocol.

The assembly protocols have been named to facilitate cross-referencing with Supp. Fig. 7.

KOD-low: For each antibody, 100 nL of the pre-assembly reaction that has undergone the 15 thermal cycles but on which the final 72 °C extension had not been performed was amplified in a 50 μL *KOD1* reaction using the appropriate construction primers (skpp101F-skpp142F and skpp101R-skpp142R).

KOD-high: For each antibody, 2 μL the full pre-assembly reaction was amplified in a 50 μL *KOD1* reaction using the appropriate construction primers (skpp101F-skpp142F and skpp101R-skpp142R).

KODXL65 and KODXL60: For each antibody, 100 nL the assembly reaction was amplified in 20 μL *KODXL* reactions using the appropriate forward and reverse construction primers. KODXL65 followed to the *KODXL* protocol exactly (with an annealing temperature of 65 °C), while KODXL60 used a 60 °C annealing temperature instead.

Nature Biotechnology: doi:10.1038/nbt.1716

Phusion72, Phusion67, and Phusion62: For each antibody, 100 nL the assembly reaction was amplified in 20 μL *Phu6* reactions with the appropriate forward and reverse construction primers. Phusion62 followed the *Phu6* protocol exactly (using an annealing temperature of 62 °C), while Phusion72 and Phusion67 used annealing temperatures of 72 °C and 67 °C, respectively.

Phusion67B, and Phusion62B: For each antibody, 100 nL the assembly reaction was amplified in 20 μL *Phu6B* reactions with the appropriate forward and reverse construction primers. Phusion62B followed the *Phu6B* protocol exactly (with the annealing temperature of 62 °C), while Phusion67B used an annealing temperature of 67 °C.

Amplification and error correction of a subset of antibodies
Based on the quality of the assemblies from the amplification optimization experiments, the following eight antibodies were chosen for cloning and characterization: efungumab, ibalizumab, panobacumab, ustekinumab, afutuzumab, oportuzumab, robatumumab, and vedolizumab. 10 nL of each pre-assemby were assembled in two 50 μL reactions following the *Phu6B* protocol using the appropriate forward and reverse primers. The reactions were cleaned up using a QIAquick PCR Purification Kit.

Error correction using ErrASE was performed as follows. 2 μL of each of the eight antibodies chosen were run a 2% E-Gel EX (Invitrogen) and reamplified by gel-stab PCR. Specifically, a 10 μL pipette tip was used to puncture the gel at the location of the desired product. The stab was mixed up and down in 10 μL of water, and the water was heated to 65 °C for 2 min. 2.5 μL of the gel-isolated product diluted in water was then amplified in a 50 μL *Phu6B* reaction. The following amount of the 8 antibody products were added to separate reactions consisting of KOD polymerase buffer (EMD chemicals, Gibbstown, NJ) containing 200 μM NTPs (EMD chemicals, Gibbstown, NJ) and 1.46 μM $MgSO_4$: 920 ng of efungumab, 630 ng of ibalizumab, 190 ng of panobacumab, 910 ng of ustekinumab, 210 ng of afutuzumab, 360 ng of oportuzumab, 420 ng of robatumumab, and 910 ng of vedolizumab. Each reaction was heated to 98°C for 1 min, cooled to 0 °C for 5 min, kept at 37 °C for 5 min, and subsequently stored and handled at 4 °C. 10 μL of the reaction was added to each of the six ErrASE reactions, and the mix was incubated at 25 °C for 1 hourf. The ErrASE reactions were then re-amplified by adding 2.5 μL of each ErrASE reaction to a 50 μL *Phu7B* reaction which used the appropriate construction primers.

Cloning

Nature Biotechnology: doi:10.1038/nbt.1716

The ErrASE-treated antibody assemblies were cleaned up using a QIAquick PCR Cleanup Kit, with the DNA eluted into 30 μL EB buffer. The 30 μL of DNA was then digested in a 100 μL reaction in FastDigest Buffer (Fermentas, Burligton, ON, Canada) that contained 4 μL of FastDigest ApaI (Fermentas) and 6 μL of FastDigest SfiI (Fermentas). The reaction was kept first at 37 °C for 30 min, and then at 50 °C for 1 h. The reactions were shaken at 800 rpm using a Thermomixer R during both thermal steps. 50 μg of the expression plasmid pSecTag2A (Invitrogen) was digested in a 100 μL of ApaI/SfiI digest similar to the one used to digest the antibody assemblies. Both the digested constructs and the digested plasmid were gel-isolated from a 2% agarose gel using a MinElute Gel Extraction Kit.

140-200 ng of one of the eight digested constructs and 90 ng of the digested plasmid were ligated in a 10 μL T4 ligase reaction the products of which were electroporated into NEB 5-alpha cells following the protocol described in the cloning of the first OLS Chip 1 constructs (with the following change: the 65 °C heat inactivation of the ligation was performed for only 10 min). The electroporated cells were suspended in 1 mL 2xYT medium, incubated at 37 °C for 45 min, and grown overnight on 50 μg/mL carbenicillin LB agar plates.

Sequencing
After a night of growth, the plates with the cloned products were sent to GENEWIZ (South Plainfield, NJ) for dideoxy sequencing. The following primers were used: forward: CMV-fwd (5'CGCAAATGGGCGGTAGGCGTG); reverse: BGHR (5'TAGAAGGCACAGTCGAGG). The trace files were analyzed using Lasergene 8[22]. Deletions of more than 2 consecutive bases were counted as single errors. Clones that had errors in >50% of the sequence were counted as misassemblies. Clones that did not have full sequence coverage between the two reads or that had traces that indicated that multiple clones were sequenced in the same reaction were counted as bad reads.

Antibodies from the second OLS chip – second set of assemblies
Amplification and processing of antibody assembly pools
Plate-specific assembly pools were amplified from the full set of 12,998 OLS2 oligos in 50 μL *Phu4* reactions with 1 μL OLS and using the plate-specific amplification primers skpp2F and skpp2R.

To make antibody assembly subpools, 20 nL of the plate subpool was amplified in 100 μL reactions following the *Phu5* protocol and using the appropriate forward and reverse amplification primers (skpp301F-skpp342F and skpp301R-skpp342R). The reaction was cleaned up using a QIAquick PCR Purification Kit, with four reactions concentrated into 120 μL EB buffer.

Nature Biotechnology: doi:10.1038/nbt.1716

119 μL (2.2 – 15.9 μg) of the antibody assembly subpools were digested with BtsI in 129 μL reactions with 0.3x NEBuffer 4, 39 ng/μL bovine serum albumin (New England Biolabs), and 0.12 units/μL BtsI (New England Biolabs). The digest was performed at 55 °C at 2 h while shaking at 1,000 rpm in the Thermomixer R. Each reaction was cleaned up using a MinElute PCR Purification Kit, with an elution into 15 of μL EB buffer. The resulting DNA concentrations ranged between 65 and 465 ng/μL, and were subsequently normalized to 50 ng/μL by adding EB buffer.

Assembly
400 ng of each antibody assembly subpool were pre-assembled in separate 20 μL reactions following the *KODpre* protocol (except without the final 5 min 72 °C extension). 10 nL of each pre-assembly reaction was then assembled into full-length genes using 50 μL *Phu7B* reactions (except that the 72 °C step during cycling was extended to 20 s) with the appropriate construction primers. Each pre-assembly was assembled in four separate reactions which were subsequently pooled. 185 μL of the assemblies were cleaned up using the QIAquick 96 PCR Purification Kit (QIAGEN), eluting into 60 μL EB with a final yield of 10-80 ng/μL.

The two antibodies that did not result in strong bands of the correct size (alacizumab and otelixizumab) were gel-stab isolated and re-amplified as follows. 20 μL of each antibody was run on a 2% E-Gel EX. A 10 μL pipette tip was used to puncture the gel at the location of the desired product. The stab was mixed up and down in 10 μL of water, and the water was heated to 60 °C for 5-20 min while being shaken at 750-800 rpm by the Thermomixer R. 1 μL the water containing the gel-isolated assemblies was then amplified in a 20 μL *Phu8B* reaction.

Error correction
Error correction using ErrASE was performed as described previously. In brief, 400 ng of abagovomab, 520 ng of alemtuzumab, 670 ng of cetuximab, 610 ng of efungumab, 310 ng of pertuzumab, 640 ng of ranibizumab, 240 ng of tadocizumab, or 660 ng of trastuzumab assembly were added to separate reactions consisting of HF Phusion buffer with 200 μM of each dNTP (Enzymatics) and either 1.5 M or no betaine (USB) (except for trastuzumab, which was error corrected only in a reaction lacking betaine) . Each reaction was heated to 98°C for 1 min, cooled to 0 °C for 5 min, kept at 37 °C for 5 min, and subsequently stored and handled at 4 °C. 10 μL of the reaction was added to each of the six ErrASE reactions, and the mix was incubated at 25 °C for 1 hour. The ErrASE reactions were then re-amplified by adding 2 μL of

Nature Biotechnology: doi:10.1038/nbt.1716

each ErrASE reaction to a 50 μL *Phu8B* reaction that used the appropriate construction primers.

Cloning
10 μg of pSecTag2A was digested in a 50 μL reaction in NEBuffer 4 with 100 ng/μL bovine serum albumin (NEB) and 2 units/μL ApaI (NEB). The digest was done for 1 h at 25 °C with shaking at 800 rpm by the Thermomixer R. At the conclusion, 2.5 μL (50 units) of SfiI (NEB) were added, and another digest was performed for 1 h at 50 °C with shaking at 800 rpm. 0.4 μL (2 units) of Antarctic phosphatase (NEB) and 5 μL of Antarctic phosphatase buffer were then added, and the reaction was allowed to proceed at 37 °C for 1 h with 800 rpm shaking. The enzymes were inactivated by heating to 70 °C for 5 min while shaking at 800 rpm.

The best ErrASE reactions were cleaned up using a QIAquick PCR Cleanup Kit, with the DNA eluted into 30 μL EB buffer. 29 μL (0.15-1.95 μg) of each assembly were digested in 50 μL reactions with NEBuffer, 100 ng/μL bovine serum albumin (NEB), and 0.8 units/μL ApaI (NEB). After 1 h at 25 °C with 800 rpm shaking, 0.5 μL (10 units) of SfiI were added and the reaction was completed with 1 h at 50 °C with 800 rpm shaking.

Both the digested constructs and the digested plasmid were gel-isolated from a 2% agarose gel using a MinElute Gel Extraction Kit.

60-175 ng of each of the digested constructs and 25 ng of the digested plasmid were ligated in a 10 μL T4 ligase reaction the products of which were electroporated into NEB 5-alpha cells following the protocol described in the cloning of the first OLS Chip 1 constructs. The electroporated cells were suspended in 1 mL EB medium, incubated at 37 °C for 70 min, and grown overnight on 50 μg/mL carbenicillin LB agar plates. Clones were picked, sequenced, and analyzed as described in the cloning of the first set of antibody assemblies from the second OLS chip.

Supplementary Table of PCR methods

Name	Buffer	Polymerase	Primers	dNTPs	Other Components	Thermocycling
KAPAprep	1x KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems, Woburn MA)	Included in Master Mix	500 nM each	Included in Master Mix		95 °C – 1 min cycle till plateau: (95 °C - 10 s 62 °C – 30 s) using BioRad CFX96 (Bio-Rad Laboratories, Hercules CA)

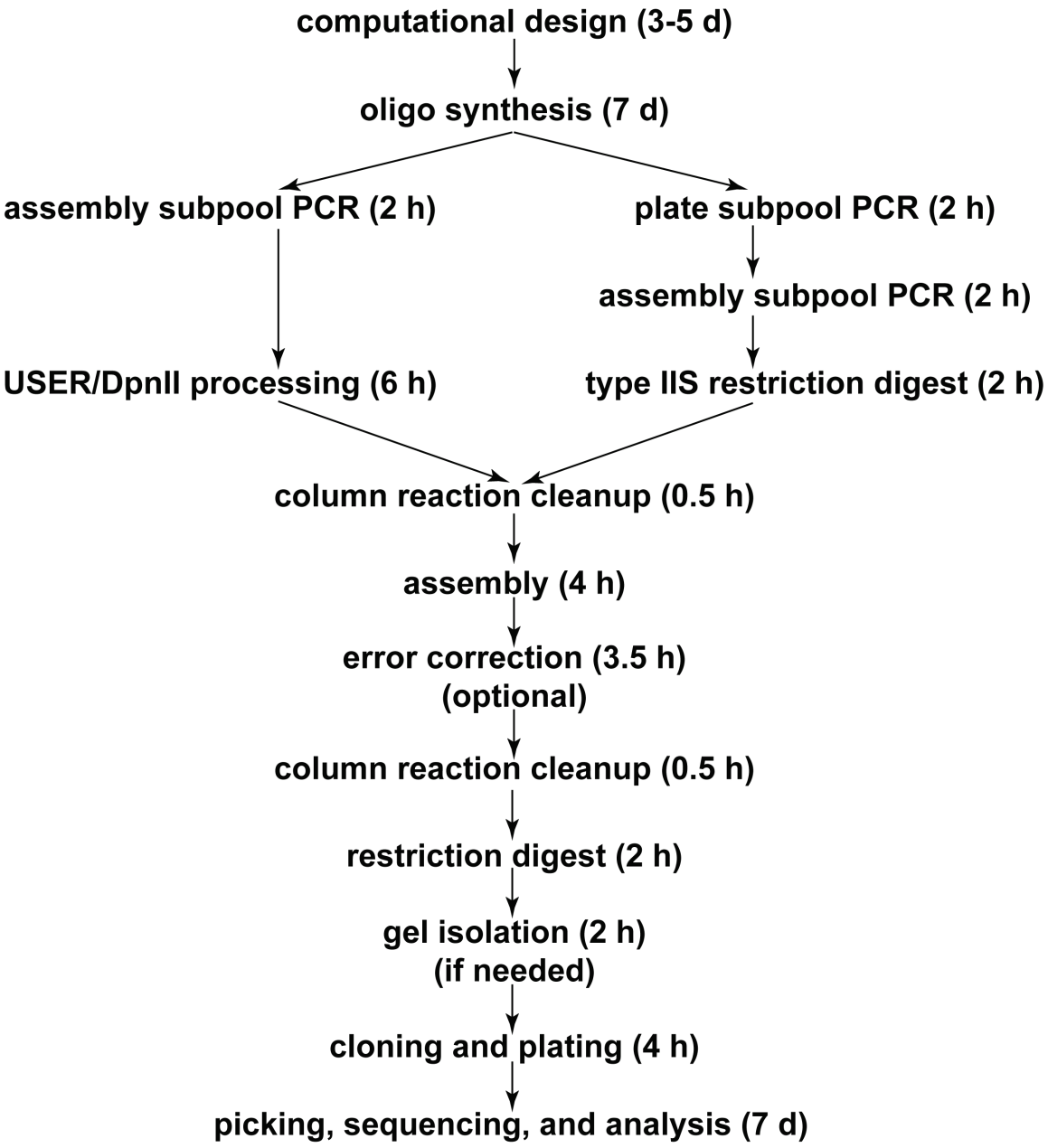
Nature Biotechnology: doi:10.1038/nbt.1716

TaqPrep	1x Taq Polymerase (Enzymatics, Beverly MA)	0.02 U/μL Taq (Enzymatics)	500 nM each	200 μM each (Enzymatics)		94 °C – 3 min 35 cycles of: (94 °C - 10 s 62 °C – 60 s) 72 – 5 min using DNA Engine Tetrad 2 (Bio-Rad)
Phu1	1x Phusion HF (Finnzymes, Woburn, MA)	0.02 U/μL Phusion (Finnzymes)	500 nM each	200 μM each (Enzymatics)		98 °C – 30 s 30 cycles of: (98 °C - 5 s 51 °C – 10 s 72 °C – 30 s) 72 – 10 min using Tetrad 2
Phu2	1x Phusion HF	0.02 U/μL Phusion	500 nM each	200 μM each (Enzymatics)		98 °C – 30 s 30 cycles of: (98 °C - 5 s 72 °C – 30 s) 72 – 10 min using Tetrad 2
Phu3	1x Phusion HF	0.02 U/μL Phusion	250 nM each	200 μM each (Enzymatics)		98 °C – 30 s 30 cycles of: (98 °C - 5 s 72 °C – 30 s) 72 – 5 min using Tetrad 2
Phu4	1x Phusion HF	0.02 U/μL Phusion	500 nM each	200 μM each (Enzymatics)		98 °C – 30 s 25 cycles of: (98 °C - 5 s 65 °C – 10 s 72 °C – 10 s) 72 – 5 min using Tetrad 2
Phu5	1x Phusion HF	0.02 U/μL Phusion	1 μM each	200 μM (Enzymatics)		98 °C – 30 s 30 cycles of: (98 °C - 5 s 65 °C – 10 s 72 °C – 10 s) 72 – 5 min using Tetrad 2
Phu6	1x Phusion HF	0.02 U/μL Phusion	500 nM each	200 μM each (Enzymatics)		98 °C – 30 s 25 cycles of: (98 °C - 5 s 62 °C – 5 s 72 °C – 10 s) 72 – 10 min using Tetrad 2
Phu6B	1x Phusion HF	0.02 U/μL Phusion	500 nM each	200 μM each (Enzymatics)	2 M betaine (USB, Cleveland OH)	98 °C – 30 s 25 cycles of: (98 °C - 5 s 62 °C – 5 s 72 °C – 10 s) 72 – 10 min using Tetrad 2
Phu7B	1x Phusion HF	0.02 U/μL Phusion	500 nM each	200 μM each (Enzymatics)	2 M betaine (USB)	98 °C – 30 s 25 cycles of: (98 °C - 5 s 62 °C – 10 s 72 °C – 15 s) 72 – 5 min using Tetrad 2

Nature Biotechnology: doi:10.1038/nbt.1716

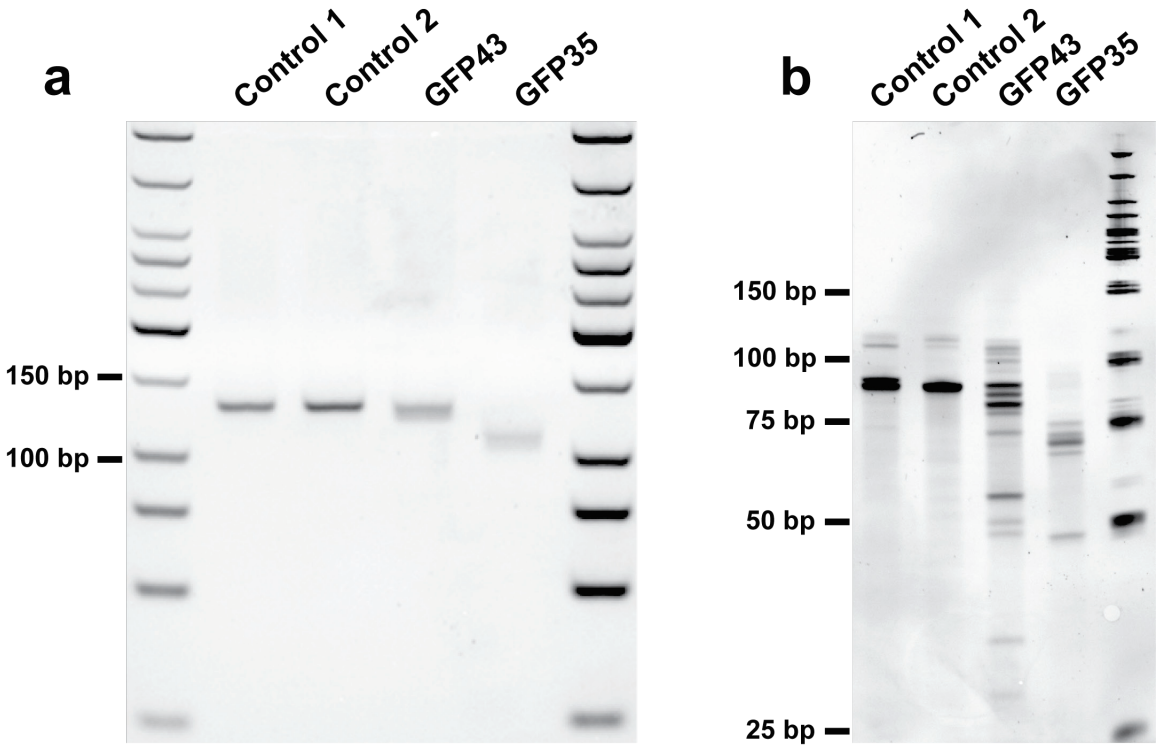
Phu8B	1x Phusion HF	0.02 U/μL Phusion	500 nM each	200 μM each (Enzymatics)	2 M betaine (USB)	98 °C – 30 s 30 cycles of: (98 °C - 5 s 62 °C – 10 s 72 °C – 20 s) 72 – 5 min using Tetrad 2
KODpre	1x KOD Polymerase (EMD Chemicals, Gibbstown NJ)	0.02 U/μL KOD (EMD Chemicals)		200 μM each (EMD Chemicals)	1.5 mM $MgSO_4$ (EMD Chemicals)	95 °C – 2 min 15 cycles of: (95 °C - 20 s 70 °C – 1 s 0.5 °C/s ramp to 50 °C 50 °C – 30 s 72 °C – 20 s) 72 – 5 min using Tetrad 2
KOD1	1x KOD Polymerase	0.02 U/μL KOD	200 nM each	200 μM each (EMD Chemicals)	1.5 mM $MgSO_4$ (EMD Chemicals)	95 °C – 2 min 25 cycles of: (95 °C - 20 s 60 °C – 30 s 72 °C – 20 s) 72 – 5 min using Tetrad 2
KODXL	KOD XL Polymerase (EMB Chemicals)	0.05 U/μL KOD XL (EMB Chemicals)	400 nM	200 μM each (EMD Chemicals)		94 °C – 30 s 25 cycles of: (94 °C - 20 s 65 °C – 5 s 74 °C – 30 s) 74 – 10 min using Tetrad 2

Nature Biotechnology: doi:10.1038/nbt.1716



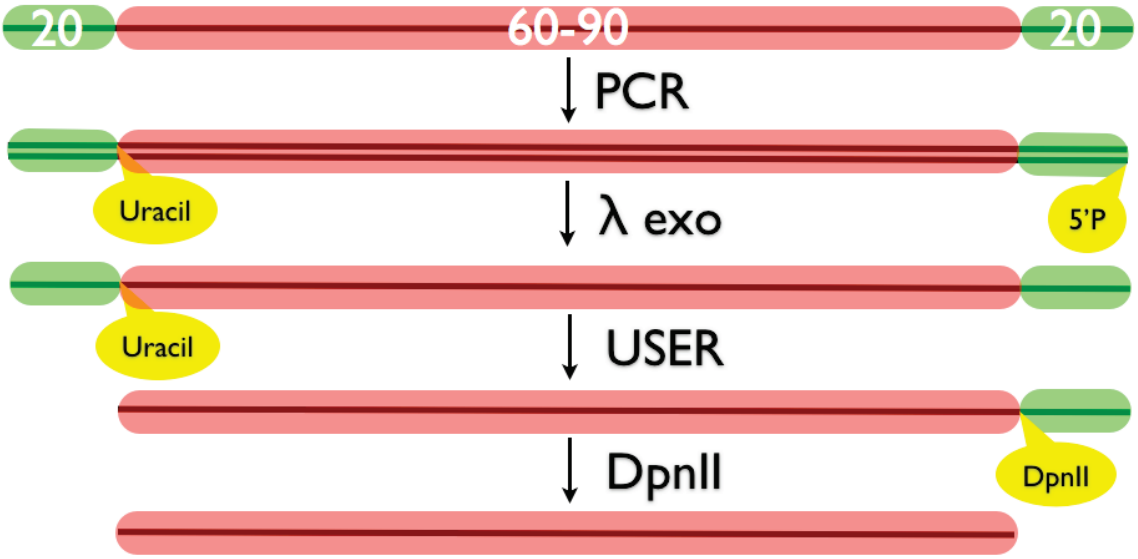
Supplementary Figure 1. Workflow for gene synthesis from high-fidelity DNA microchips. Shown here are the major steps and approximate timings of the entire gene synthesis process. The branch point reflects the choice of whether USER/DpnII processing (left branch after oligo synthesis) or type IIS enzymatic processing (right branch) are used for removing the amplification sites. The process outlines the final optimized form of the optimized protocols. The times given in parentheses are estimates that account for the time involved in both setting up and running the reactions.

Nature Biotechnology: doi:10.1038/nbt.1716



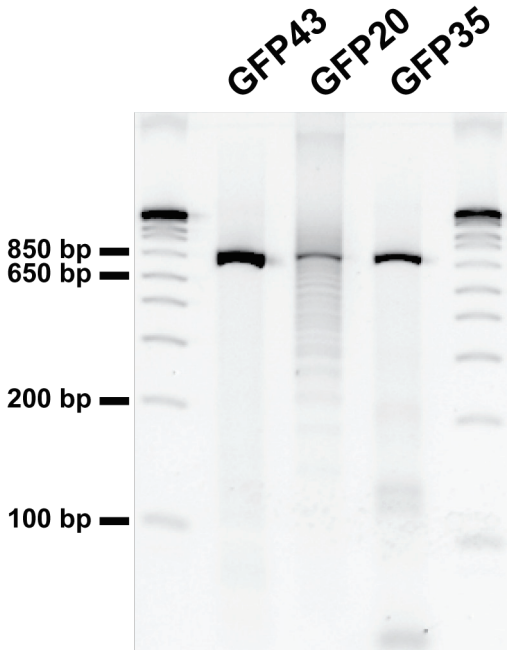
Supplementary Figure 2. Amplification and processing of OLS Pool 1 oligonucleotides. Two assembly subpools and two control subpools were amplified from OLS Pool 1, which contained a total of 13,000 oligonucleotides (**a**). Because the oligonucleotides in the two GFP subpools had sizes distinct from all other oligonucletides on the chip, and since no oligonucleotides of the wrong size were detected, these data suggest that the oligonucleotides from other subpools did not amplify. The dsDNA subpools were then processed using DpnII/USER/lambda exonuclease (**b**). After processing there were three types of products. First, there were the products of the expected size. Second, there were the high molecular weight products that correspond to oligonucleotides that retain one or both of the assembly-specific primer sites. Last, there were the low molecular weight products that were likely produced by DpnII cleavage at double-stranded recognition sites formed by the overlapping regions of the oligonucleotides. The same dsDNA ladder (Low Molecular Weight, NEB) is used in both the non-denaturing (**a**) and the denaturing (**b**) 10% PAGE gels, where the denaturing agent produced the extra bands in the ladder in (**b**).

Nature Biotechnology: doi:10.1038/nbt.1716



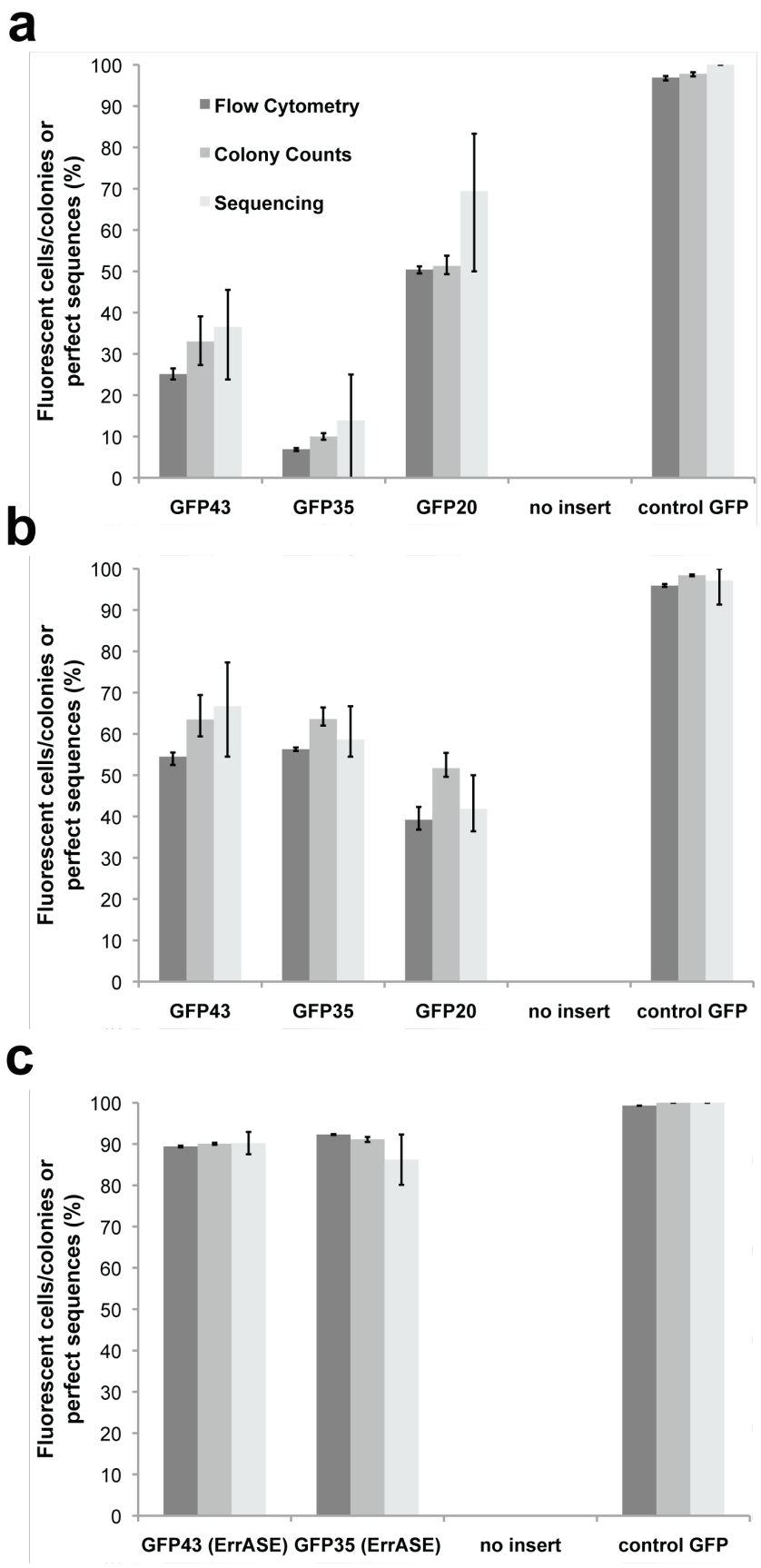
Supplementary Figure 3. OLS Pool 1 assembly subpool amplification and USER/DpnII processing. Assembly subpools were amplified from OLS Pool 1 using 20 bp priming sites that were shared by all primers in any particular assembly. A PCR reaction resulted in a pool of dsDNA with the oligos from other assemblies still in ssDNA form and at trace concentrations. The forward subpool amplification primers incorporates two sequential phosphorothioate linkages on the 5' end, and a deoxyuridine its 3' end, while the reverse primer had a phosphate at its 5' end. Lambda exonuclease is a 5' -> 3' exonuclease that degrades 5' phosphorylated DNA and is blocked by phoshporothioate. This property was used to selectively degrade the remove strand of the amplified products. USER (Uracil-Specific Excision Reagent) Enzyme (NEB) removed the 5' priming site by excising the uracil and cutting 3' and 5' of the resulting apyrimidinic site. Next, the 3' end was annealed to the reverse amplification primer, forming a double-stranded DpnII recognition site (5'GATC). The 3' priming site was then removed with a DpnII digest.

Nature Biotechnology: doi:10.1038/nbt.1716



Supplementary Figure 4. First GFP assembly from OLS Pool 1. The two OLS Pool 1 GFP assembly subpools were amplified, processed, and PCR assembled (Figure 3a). The bands corresponding to full-length assembly products were then gel-isolated and re-amplified. The re-amplification products shown above contained low molecular weight products that likely remained in trace amounts after gel isolation. These significantly increased the number of clones that needed to be sequenced in order to find a perfect GFPmut3 construct. The control GFP was amplified from a cloned GFP. GFP20 was an assembly made from a hand-mixed pool of oligonucleotides synthesized using a column-based method. GFP20 was not gel isolated or re-amplified.

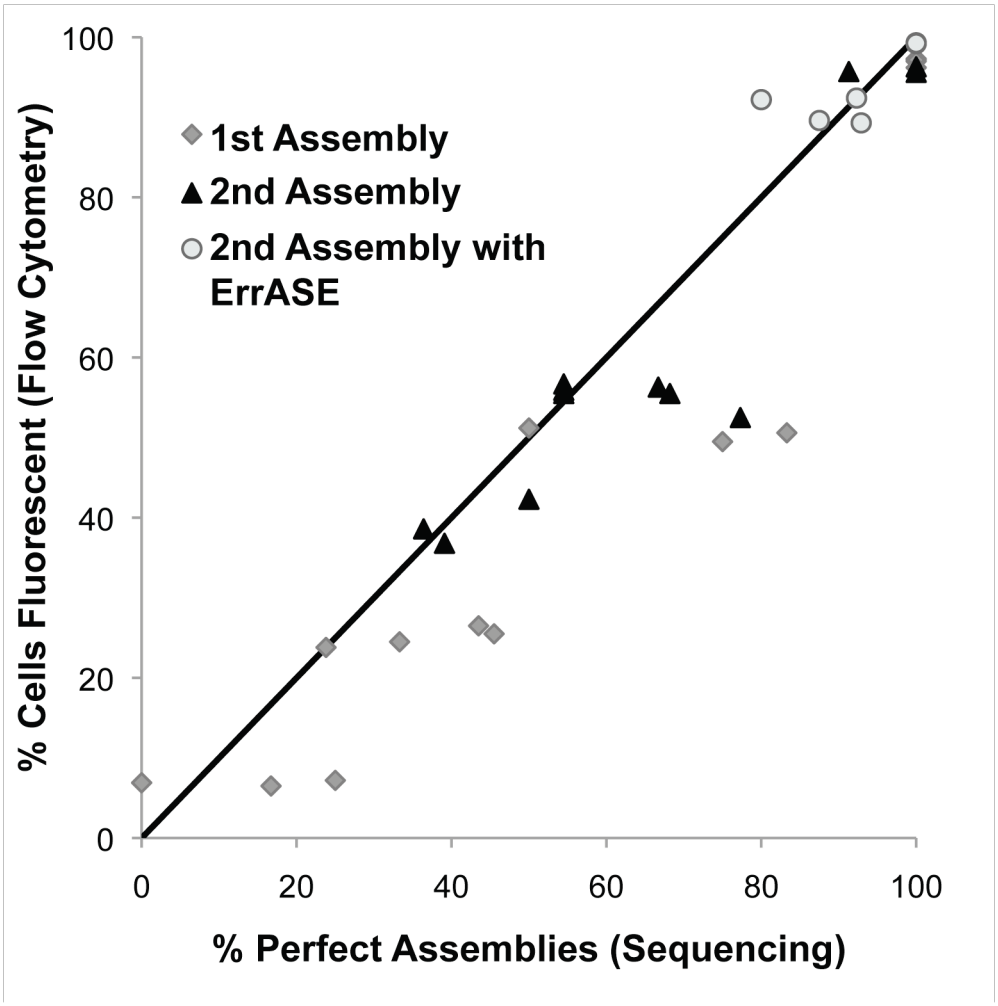
Nature Biotechnology: doi:10.1038/nbt.1716



Nature Biotechnology: doi:10.1038/nbt.1716

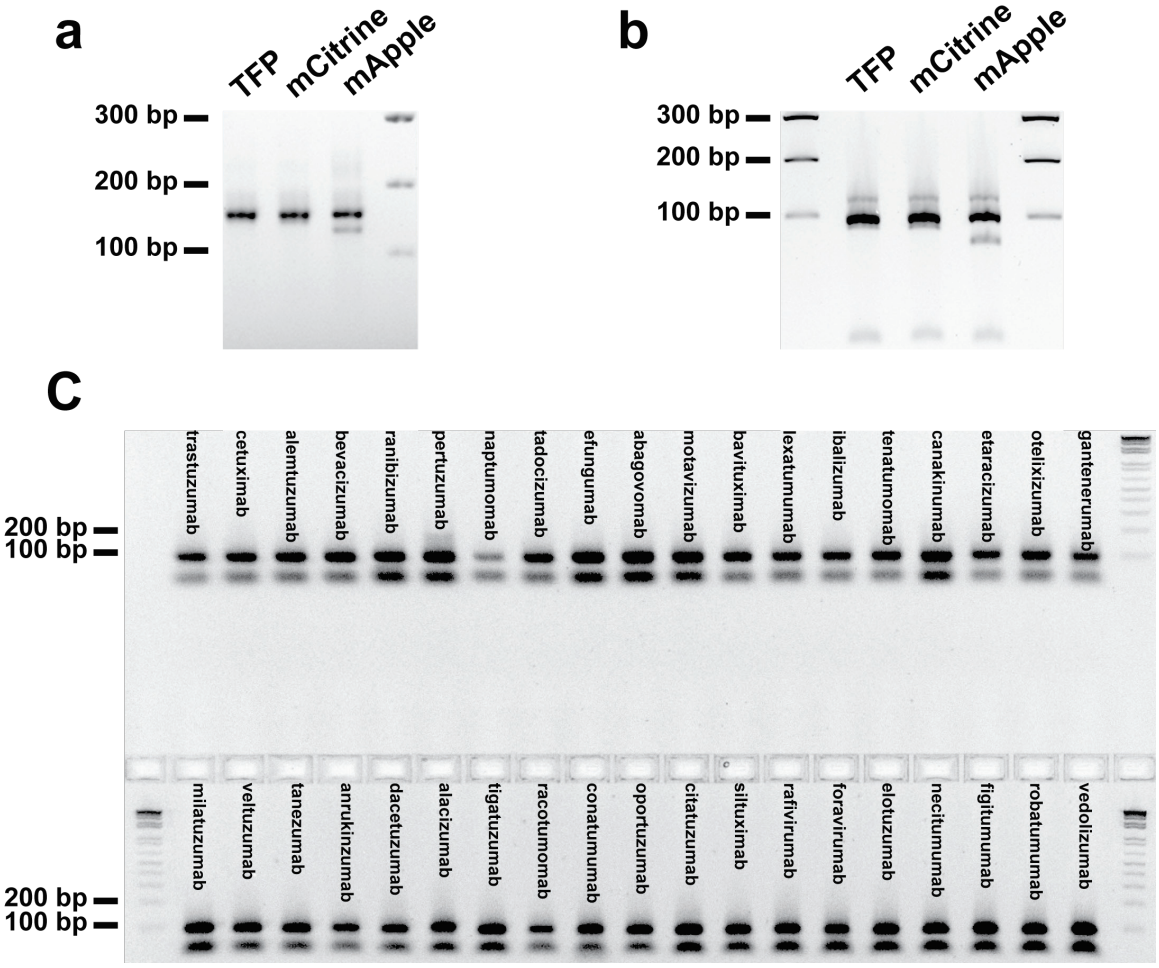
Supplementary Figure 5. Screening error rates of GFP assemblies. Error rates from the first set (gel-isolated and re-amplified) (**a**), the second set (gel-isolated without re-amplification) (**b**), and the error-corrected second set of GFP assemblies from OLS Pool 1 were determined using flow cytometry, by counting colonies on agar plates, and by sequencing individual clones. Error bars give the range of the observed values. n corresponds to the number of electroporated cultures from one or more ligation reactions performed on a single assembly reaction, with n = 3-4 in (**a**) n = 3 in (**b**), and n = 2 in (**c**).

Nature Biotechnology: doi:10.1038/nbt.1716



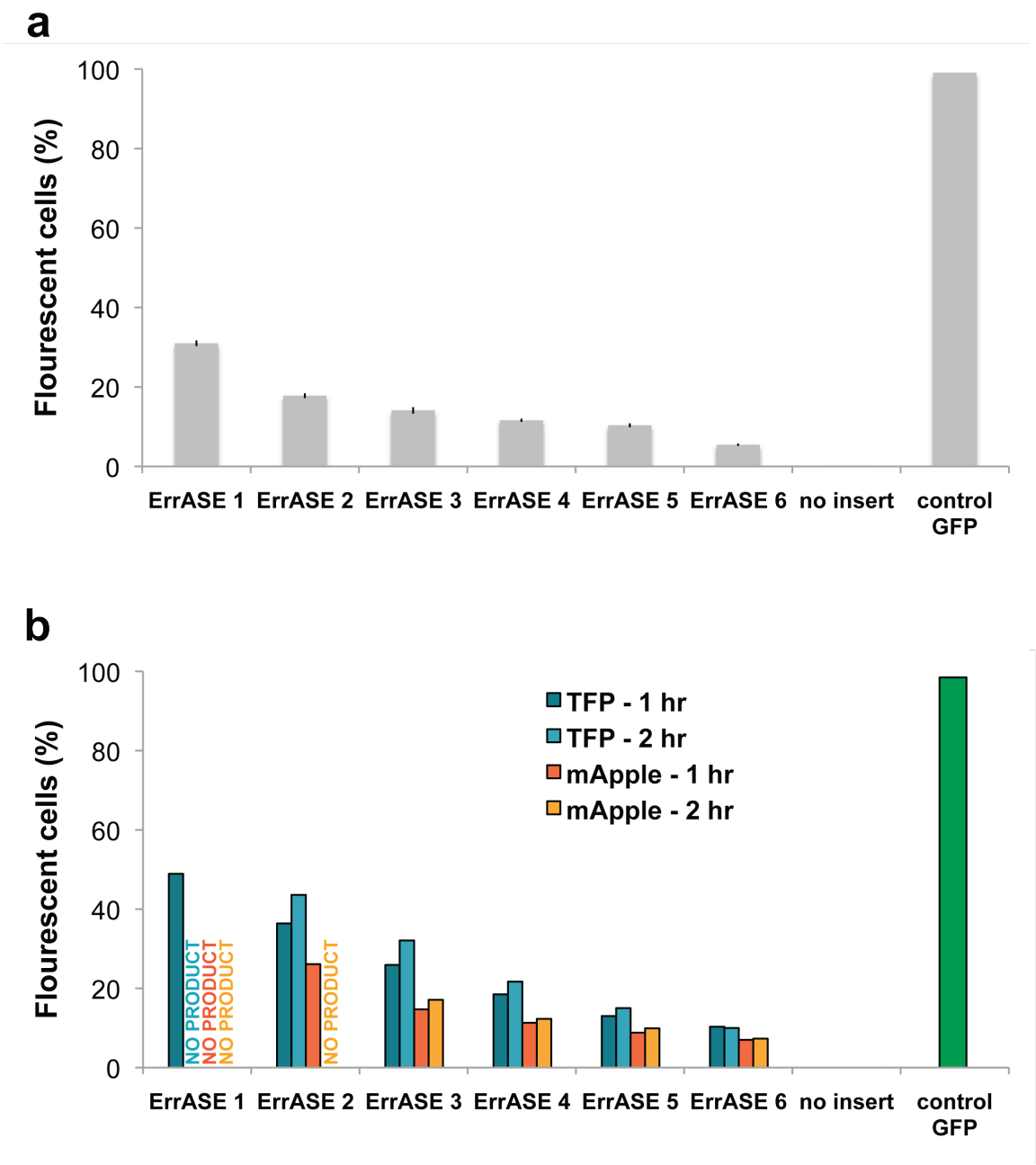
Supplementary Figure 6. The dynamic range of the flow cytometry screen. The relationship between the fluorescent fraction observed with flow cytometry is shown as a function of the fraction of perfect assemblies predicted from the sequencing data, with each data point corresponding to individual samples constructs built from the OLS Pool 1 (the same data are shown in Supp. Fig 5). The black line indicates the result expected if the sequencing and fluorescent data predicted each other perfectly.

Nature Biotechnology: doi:10.1038/nbt.1716



Supplementary Figure 7. Processing OLS Pool 2 assembly subpools. Assembly-specific primers were used to amplify the subpools that were designed to build three different fluorescent proteins (**a**). A BtsI restriction enzyme removed the priming sites (**b**). The same protocol was followed in processing the antibody assembly subpools, with (**c**) showing the subpools after the BtsI digest. The gel in (**c**) shows only 38 subpools because other 4 antibody subpools evaporated from the reaction tubes during PCR and had to be re-amplified in a separate experiment.

Nature Biotechnology: doi:10.1038/nbt.1716

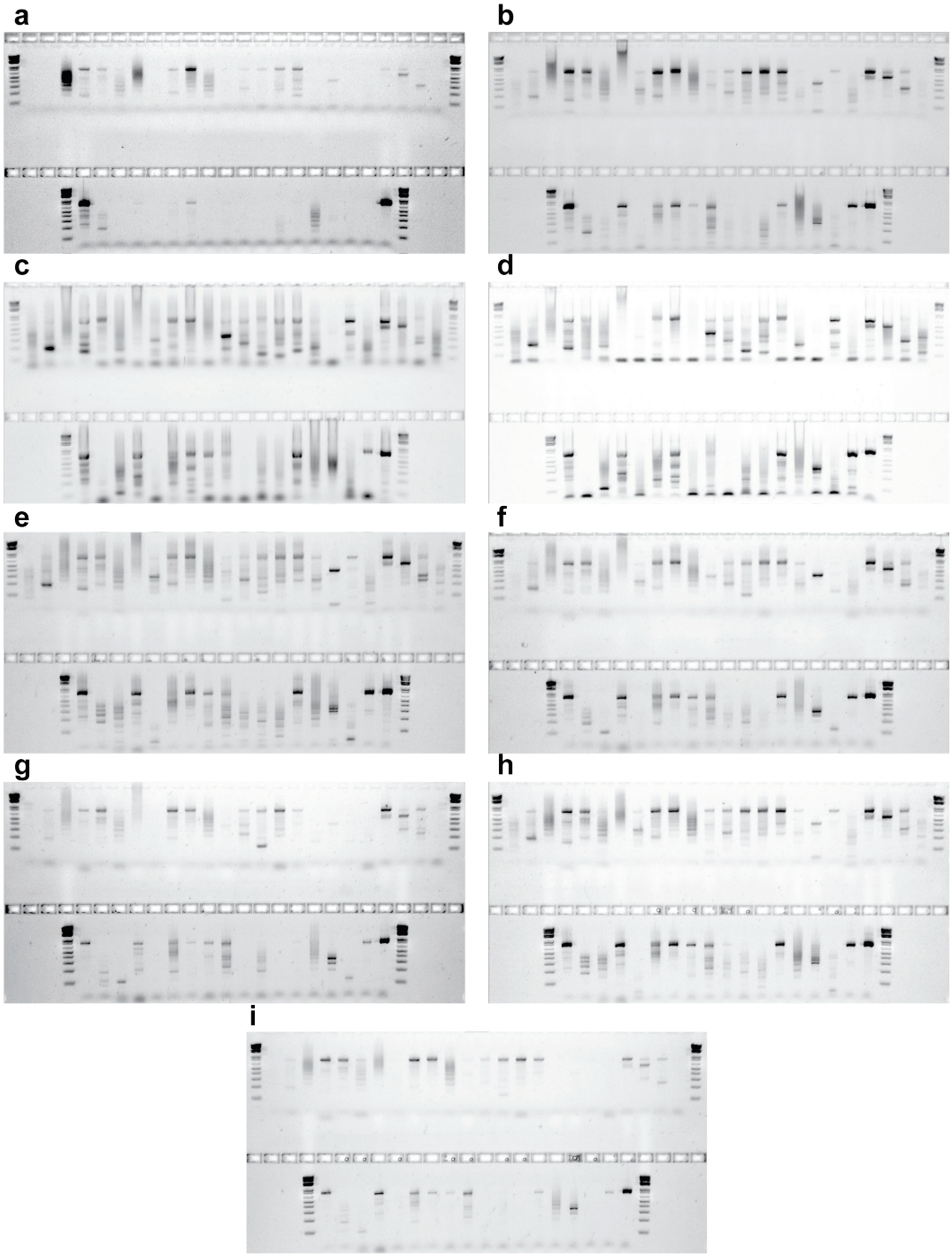


Supplementary Figure 8. Optimization of enzymatic synthesis error removal with ErrASE. mCitrine synthesized from OLS Pool 2 was treated with ErrASE and the fluorescent fraction was quantified with flow cytometry (**a**). The different ErrASE reactions correspond to decreasing stringency of ErrASE that were provided by the manufacturer, with ErrASE 1 having the most stringency and ErrASE 6 having the least. Error bars give the range of the data points, with n = 2 or 4 for the control and the mCitrine constructs, respectively. Increasing the length of ErrASE treatment from 1 to 2 h did not lead to a major decrease in error rates (**b**). "NO PRODUCT" indicates that the post-ErrASE amplification did not product a product of the right size, most likely because the ErrASE error removing enzymes have over-digested

Nature Biotechnology: doi:10.1038/nbt.1716

the assembly. Each value is an average of independent flow cytometry runs performed on 5 (**a**) or 3 (**b**) aliquots of the cloned assemblies.

Nature Biotechnology: doi:10.1038/nbt.1716

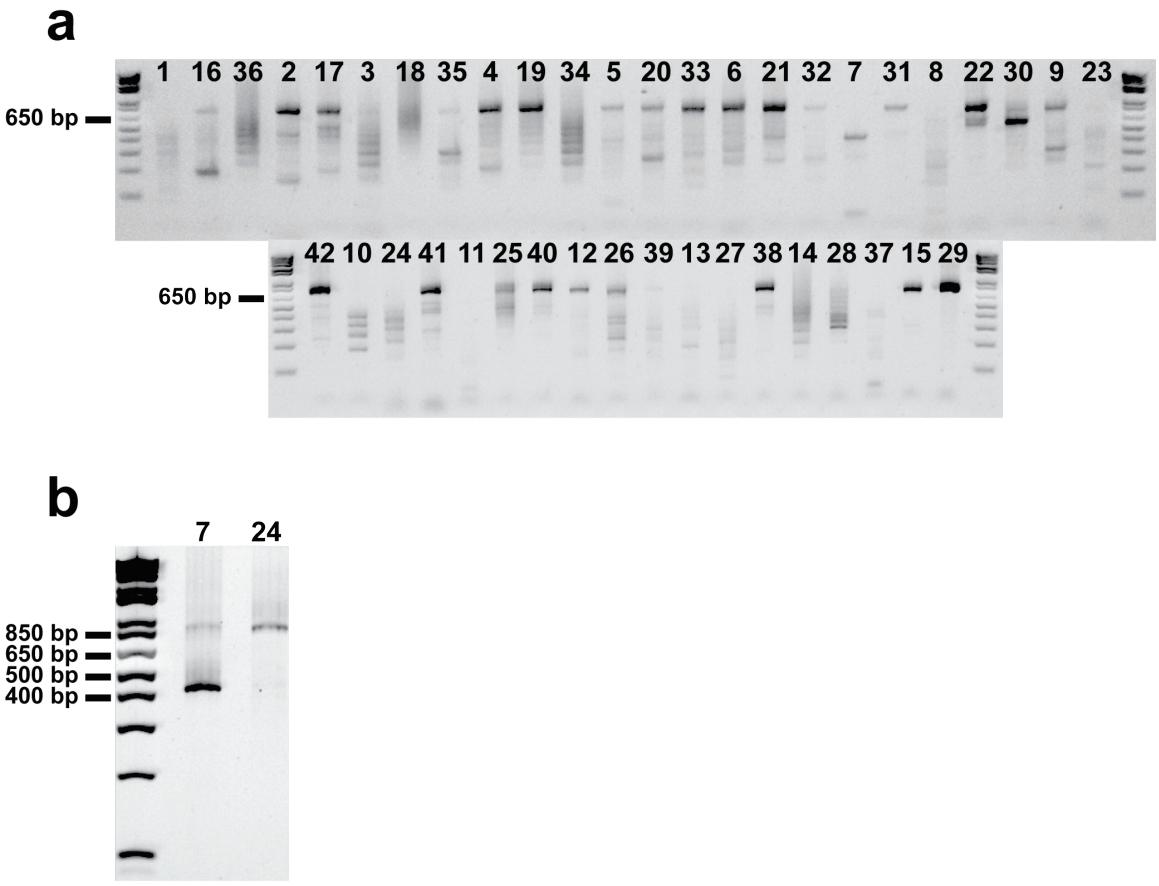


Supplementary Figure 9. Optimizing the antibody assembly protocol. First, each antibody assembly subpool was subjected to 15 PCR cycles in the presence of KOD DNA polymerase but in the absence of the construction primers. Next, the construction primers, and each assembly was diluted in

Nature Biotechnology: doi:10.1038/nbt.1716

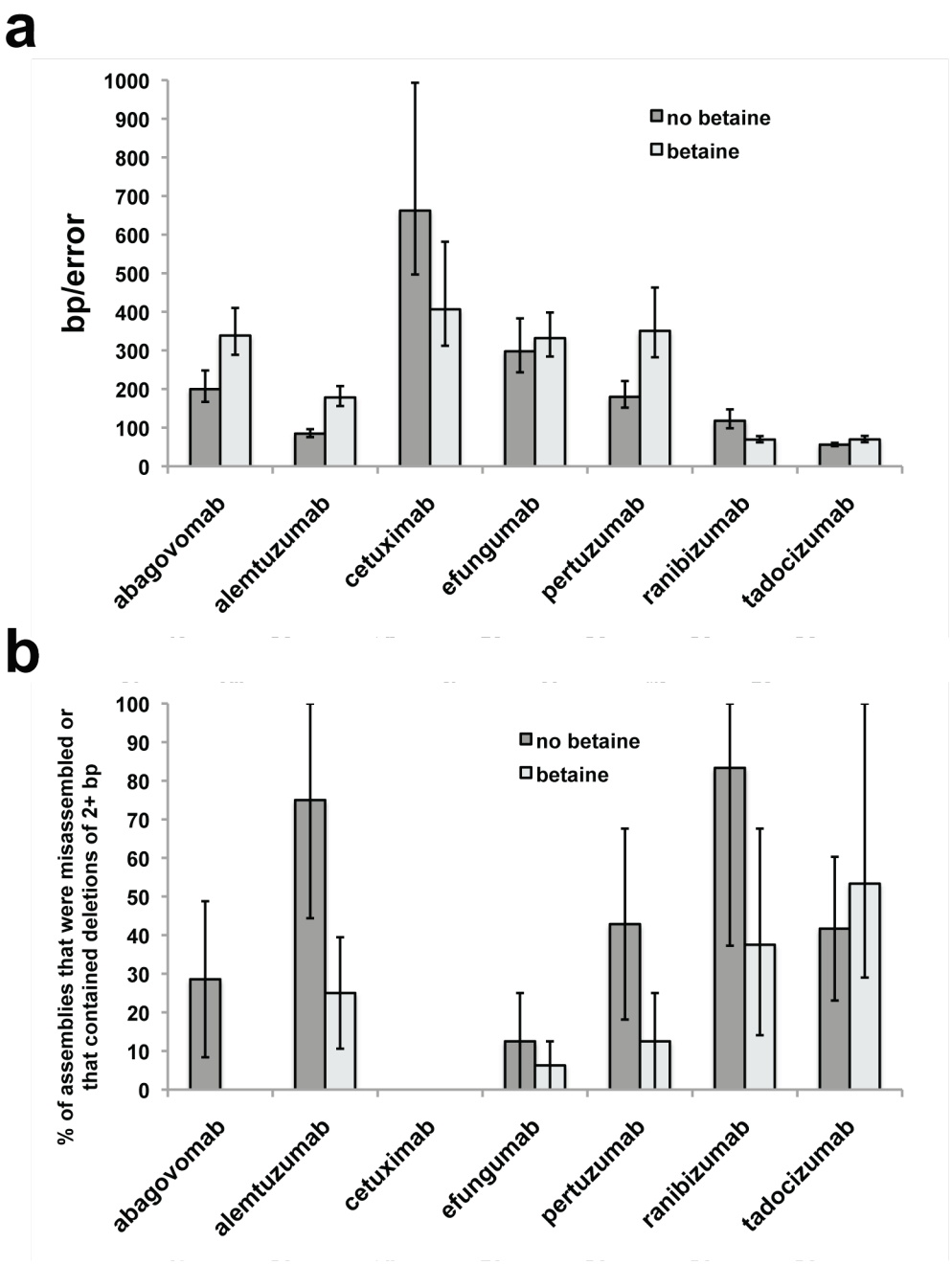
another PCR mix. Show are the 2% agarose gels of the following assembly protocols (as described in the Supp. Methods): (**a**) KOD-low, (**b**) KOD-high, (**c**) KODXL60, (**d**) KODXL65, (**e**) Phusion62, (**f**) Phusion67, (**g**) Phusion 72, (**h**) Phusion 62B, (**i**) Phusion67B. The lane order in each lane is identical to that in Supplementary Fig. 10a. Invitrogen's 1 Kb Plus DNA Ladder was used as a size marker in all experiments.

Nature Biotechnology: doi:10.1038/nbt.1716



Supplementary Figure 10. Initial antibody assemblies. (**a**) The initial assembly reaction resulted in 29 out of 42 antibody assembly reactions yielding products of the correct size (rather than 40/42 in the final optimized assembly reactions). The antibody that corresponds to each number is listed in Supp. Table 3. Increasing the assembly subpool concentration used in the assembly reaction increased the number of successful assemblies to 40 (see Fig. 2d). The two failures from the final optimized assembly reactions were gel-isolated and re-amplified, yielding products of the correct size (**b**).

Nature Biotechnology: doi:10.1038/nbt.1716



Supplementary Figure 11. Testing the use of betaine during the ErrASE melt and re-anneal step. A set of synthesized antibodies (synthesis products shown in Fig. 2d) was error corrected using the ErrASE protocol with betaine either included or left out of the melting and re-annealing step. The resulting error rate (**a**) and the probability of a synthesized molecule being either misassembled or having a large (3+ consecutive bp) deletion (**b**) was quantified. Error bars indicate the expected Poisson error (√n, with n being the number of errors observed).

Nature Biotechnology: doi:10.1038/nbt.1716

55K SLXA	Pre-PCR OLS Pool	Post-PCR OLS Pool
Total Reads	757126	830659
Mapped reads	530616	616713
Mapped reads <34bp	14426	20982
Imperfect Oligos	67050	78769
Avg Error of Imperfect Oligo	1.67	1.69
Phred30 Imperfect Oligos	28812	29033
Phred30 Average Error of Imperfect Oligo	1.286	1.305
Matches	18466976	21454745
Transitions	24569	56377
Transversions	66905	81820
Deletions	19761	24016
Insertions	839	935
Match %	99.40%	99.25%
Transition %	0.13%	0.26%
Transversion %	0.36%	0.38%
Deletion %	0.11%	0.11%
Insertion %	0.00%	0.00%
Phred30 Matches	17443050	20217413
Phred30 Transitions	10914	8908
Phred30 Transversions	10743	10369
Phred30 Deletions	14795	17965
Phred30 Insertions	600	659
Phred30 Match %	99.79%	99.81%
Phred30 Transition %	0.06%	0.04%
Phred30 Transversion %	0.06%	0.05%
Phred30 Deletion %	0.08%	0.09%
Phred30 Insertion %	0.00%	0.00%

Supplementary Table 1. Reanalysis of Agilent OLS libraries for quantitation of error rates. We reanalyzed a previously published Illumina data set for determining sequencing error rates of 53,777 oligonucleotides produced on the OLS platform both before and after two rounds of PCR amplification[23]. *Total Reads* refer to the number of reads from a single lane of Illumina sequencing. *Mapped reads* refer to the total number of reads that mapped to one of the designed oligonucleotides. *Mapped reads <34bp* refer to reads that were only partial matches to the designed oligonucleotides, and are probably due to misalignment. *Imperfect oligos* are the number of oligos that are not perfect matches to the mapped sequence. The *Avg Error of Imperfect Oligo* are the average number of

Nature Biotechnology: doi:10.1038/nbt.1716

errors in an oligonucleotide given that it is not perfect (this is to check if there were a clustering of errors). The Phred30 versions of *Imperfect oligos* and *Avg Error of Imperfect Oligo* are the same counts, but only counting errors that pass the Phred30 quality scores for base calls. The next two sets of data give information on the types of errors detected in the oligos either filtered by quality scores of the base calls (Phred30) or not. *Matches* refer to bases that are equivalent to the designed sequences. *Transitions* and *Transversions* refer to types of mismatches from the designed sequenced. *Insertions* and *Deletions* refer to the number of single base insertions or deletions present in the sequenced oligo as compared to the designed.

Nature Biotechnology: doi:10.1038/nbt.1716

Construct	Good Reads	Missassemblies	Per-fect	Sequenced Bases	Mis-matches	Small Dele-tions	Large Dele-tions (>2bp)	Large Dele-tion Size	Inser-tions	Bp / Error	Poisson High	Poisson Low
GFP20	49	4	28	35133	0	3	0	0	6	1351	330	222
GFP43	63	1	44	45171	5	17	0	0	8	1506	336	232
GFP43 (ErrASE)	30	0	27	21510	3	0	0	0	0	7170	9794	2624
GFP35	60	0	36	43020	5	29	0	0	4	1132	219	158
GFP35 (ErrASE)	28	0	24	20076	1	3	0	0	0	5019	5019	1673
abagovomab	15	0	1	11175	20	12	0	0	1	339	71	50
afutuzumab	15	1	2	11580	24	7	0	0	0	374	82	57
alemtuzumab	12	0	0	8913	22	19	9	99	0	178	29	22
cetuximab	8	0	2	5960	6	3	0	0	0	662	331	166
efungumab	16	0	2	11945	27	8	1	23	0	332	66	47
ibalizumab	8	0	0	6224	11	2	0	0	0	479	184	104
panobacumab	22	1	3	16707	38	23	3	13	0	261	37	29
pertuzumab	8	0	3	5959	10	4	2	25	1	351	112	68
ranibizumab	4	2	0	2948	7	11	7	80	0	118	29	20
robatumumab	21	0	0	14860	36	20	24	911	2	181	22	18
tadocizumab	7	8	0	5200	43	18	1	15	13	69	9	7
trastuzumab	16	0	1	11772	24	25	10	196	1	196	29	22
ustekinumab	23	0	6	17336	32	11	1	6	0	394	70	52
vedolizumab	33	0	6	25571	43	9	1	4	0	482	77	58

Supplementary Table 2. Sequencing results for cloned assemblies. The results from sequencing of constructs made from IDT oligonucleotides (GFP20), OLS Pool 1 (GFP43 & GFP35), and OLS Pool 2 (antibodies). The data for each antibody were taken from the cloning experiment that had the lowest error rate. *Good reads* are the number of clones that returned sequence information. *Misassemblies* are the number sequences that had less than 50% of the correct sequence. *Perfect reads* are the number of clones that had the exact sequence expected. *Sequenced bases* are the total number of sequenced bases homologous to the designed sequence and mismatches to the number of mismatches from the designed sequences. *Small* and *large indels* are the number of deletions <3 or >2 bp consecutively, respectively. *Large Deletion Size* is the sum of to the cumulative number of bases deleted in the large indels. *Insertions* refer to the number of incorrectly inserted bases in the sequence. *Bp/error* refers to the average error rate, with each long deletion counted as a single error. *Poisson high* and *poisson low* are the expected Poisson noise (minus and plus the square of the number of errors, respectively). The same data are presented graphically in Fig. 3b.

Nature Biotechnology: doi:10.1038/nbt.1716

Name	ID #	Primers (subpool/construction)	Linker	Band from Assembly?	Reaction Cloned	Perfect Clone Found?
trastuzumab	1	301/101	GGSGGSGGASGAGSGGG	yes	2	yes
bevacizumab	2	304/104	GGSGGSGGASGAGSGGG	yes		
pertuzumab	3	306/106	GGSGGSGGASGAGSGGG	yes	2	yes
efungumab	4	309/109	GGSGGSGGASGAGSGGG	yes	1 and 2	yes
bavituximab	5	312/112	GGSGGSGGASGAGSGGG	yes		
tenatumomab	6	315/115	GGSGGSGGASGAGSGGG	yes		
otelixizumab	7	318/118	GGSGGSGGASGAGSGGG	no (very faint)		
gantenerumab	8	320/120	GGSGGSGGASGAGSGGG	yes		
tanezumab	9	323/123	GGSGGSGGASGAGSGGG	yes		
dacetuzumab	10	326/126	GGSGGSGGASGAGSGGG	yes		
racotumomab	11	329/129	GGSGGSGGASGAGSGGG	yes		
oportuzumab	12	332/132	GGSGGSGGASGAGSGGG	yes	1 (none sequenced)	
rafivirumab	13	335/135	GGSGGSGGASGAGSGGG	yes		
elotuzumab	14	338/138	GGSGGSGGASGAGSGGG	yes		
robatumumab	15	341/141	GGSGGSGGASGAGSGGG	yes	1	no
cetuximab	16	302/102	GGSAGSGSSGGASGSGG	yes	2	yes
ranibizumab	17	305/105	GGSAGSGSSGGASGSGG	yes	2	no
naptumomab	18	307/107	GGSAGSGSSGGASGSGG	yes		
abagovomab	19	310/110	GGSAGSGSSGGASGSGG	yes	2	yes
lexatumumab	20	313/113	GGSAGSGSSGGASGSGG	yes		
canakinumab	21	316/116	GGSAGSGSSGGASGSGG	yes		
milatuzumab	22	321/121	GGSAGSGSSGGASGSGG	yes		
anrukinzumab	23	324/124	GGSAGSGSSGGASGSGG	yes		
alacizumab	24	327/127	GGSAGSGSSGGASGSGG	no		
conatumumab	25	330/130	GGSAGSGSSGGASGSGG	yes		
citatuzumab	26	333/133	GGSAGSGSSGGASGSGG	yes		
foravirumab	27	336/136	GGSAGSGSSGGASGSGG	yes		
necitumumab	28	339/139	GGSAGSGSSGGASGSGG	yes		
vedolizumab	29	342/142	GGSAGSGSSGGASGSGG	yes	1	yes
veltuzumab	30	322/122	GGAGSGAGSGSSGAGSG	yes		
panobacumab	31	319/119	GGAGSGAGSGSSGAGSG	yes	1	yes
etaracizumab	32	317/117	GGAGSGAGSGSSGAGSG	yes		
ibalizumab	33	314/114	GGAGSGAGSGSSGAGSG	yes	1	no
motavizumab	34	311/111	GGAGSGAGSGSSGAGSG	yes		
tadocizumab	35	308/108	GGAGSGAGSGSSGAGSG	yes	2	no
alemtuzumab	36	303/103	GGAGSGAGSGSSGAGSG	yes	2	no
figitumumab	37	340/140	GGAGSGAGSGSSGAGSG	yes		
farletuzumab	38	337/137	GGAGSGAGSGSSGAGSG	yes		
siltuximab	39	334/134	GGAGSGAGSGSSGAGSG	yes		
afutuzumab	40	331/131	GGAGSGAGSGSSGAGSG	yes	1	yes
tigatuzumab	41	328/128	GGAGSGAGSGSSGAGSG	yes		
ustekinumab	42	325/125	GGAGSGAGSGSSGAGSG	yes	1	yes

Supplementary Table 3. Assembly results from 42 attempted antibody constructions. Of the 42 assemblies of antibody subpools from OLS Pool 2, 29 of the first set of reactions (Supp. Fig. 8a) and 40 of the second set (Fig. 3d) resulted in products of the correct size. We attempted

Nature Biotechnology: doi:10.1038/nbt.1716

to clone 8 from the first set of assemblies (7 cloned successfully) and 8 from the second (all cloned successfully). The *ID #* refers to the number used throughout the paper to identify the antibody. *Primers* are the primer number as given in the Supplementary Sequences, with a forward and reverse primer pair corresponding to each number (for instance, skpp-301-F and skpp-301-R are the assembly subpool amplification primers for trastuzumab). *Linker* refers to the amino acid sequence used to link the heavy and the light chain. *Band from assembly?* refers to presence of a band of the correct size refers to the gel in Fig. 2d. The *Reaction cloned* column indicates whether the antibody was cloned from either of two assembly reactions (assembly 1 shown in Supplementary Fig. 10a, assembly 2 shown in Fig. 3d). Oportuzumab was cloned, but no assembly products were found among the sequenced clones. *Perfect clone found?* Indicates whether or not at least one of the cloned assemblies sequenced contained no errors.

Nature Biotechnology: doi:10.1038/nbt.1716

Supplementary References

1. Tian, J. *et al*. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050-1054 (2004).
2. Zhou, X. *et al*. Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.* **32**, 5409-5417 (2004).
3. Richmond, K.E. *et al*. Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32**, 5011-5018 (2004).
4. Kim, C. *et al*. Progress in gene assembly from a MAS-driven DNA microarray. *Microelectronic Eng.* **83**, 1613-1616 (2006).
5. Cormack, B.P., Valdivia, R.H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38 (1996).
6. Bang, D. & Church, G.M. Gene synthesis by circular assembly amplification. *Nat. Methods* **5**, 37-39 (2008).
7. Rouillard, J. *et al*. Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Res.* **32**, W176-180 (2004).
8. Xu, Q. *et al*. Design of 240,000 orthogonal 25mer DNA barcode probes. *P. Natl. Acad. Sci. USA* **106**, 2289-2294 (2009).
9. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *P. Natl. Acad. Sci. USA* **95**, 1460-1465 (1998).
10. SantaLucia, J. & Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Bioph. Biom.* **33**, 415-440 (2004).
11. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).
12. Markham, N.R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. *Meth. Mol. Biol.* **453**, 3-31 (2008).
13. Kaplinski, L. *et al*. MultiPLX: automatic grouping and evaluation of PCR primers. *Bioinformatics* **21**, 1701-1702 (2005).
14. Cock, P.J. *et al*. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).
15. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203-1210 (1997).
16. Isaacs, F. *et al*. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841-847 (2004).
17. Vallone, P.M. & Butler, J.M. AutoDimer: a screening tool for primer-dimer and hairpin structures. *BioTechniques* **37**, 226-231 (2004).
18. Ai, H. *et al*. Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural

Nature Biotechnology: doi:10.1038/nbt.1716

characterization and applications in fluorescence imaging. *Biochem. J.* **400**, 531-540 (2006).
19. Griesbeck, O. *et al*. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J. Biol. Chem.* **276**, 29188-29194 (2001).
20. Shaner, N.C. *et al*. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545-551 (2008).
21. Lefranc, M. *et al*. IMGT(R), the international ImMunoGeneTics information system(R). *Nucleic Acids Res.* **37**, D1006-D1012 (2009).
22. Burland, T.G. DNASTAR's Lasergene sequence analysis software. *Meth. Mol. Biol.* **132**, 71-91 (1999).
23. Li JB, Gao Y, Aach J, et al. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res.* **19**, 1606-1615 (2009).

Nature Biotechnology: doi:10.1038/nbt.1716