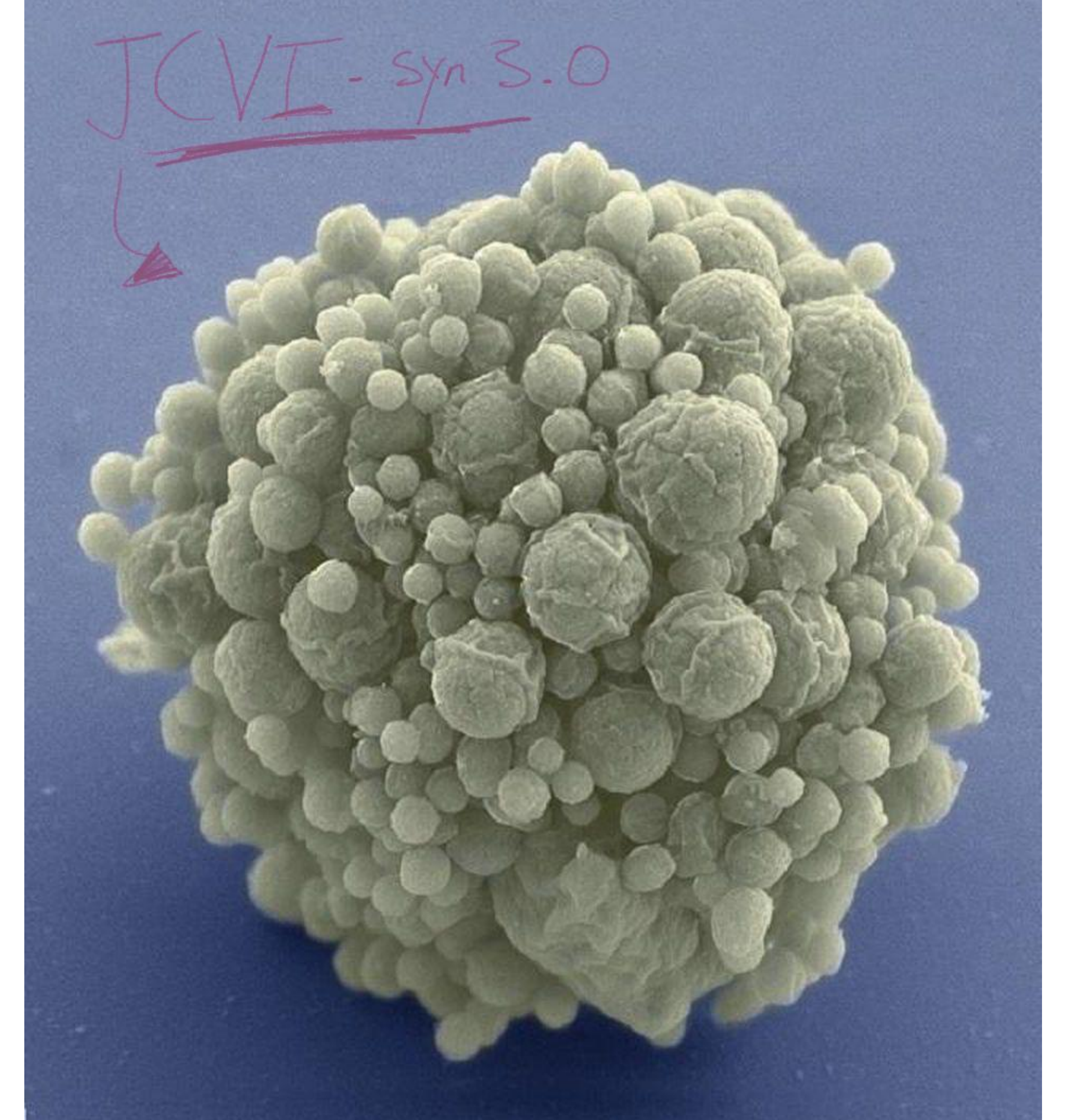

JCVI - Syn 3.0

Corrected 28 June 2016; see full text.

RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Design and synthesis of a minimal bacterial genome

Clyde A. Hutchison III,*† Ray-Yuan Chuang,† Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, Krishna Kannan, Bogumil J. Karas, Li Ma, James F. Pelletier, Zhi-Qing Qi, R. Alexander Richter, Elizabeth A. Strychalski, Lijie Sun, Yo Suzuki, Billyana Tsvetanova, Kim S. Wise, Hamilton O. Smith, John I. Glass, Chuck Merryman, Daniel G. Gibson, J. Craig Venter*

INTRODUCTION: In 1984, the simplest cells capable of autonomous growth, the mycoplasmas, were proposed as models for understanding the basic principles of life. In 1995, we reported the first complete cellular genome sequences (*Haemophilus influenza*, 1815 genes, and *Mycoplasma genitalium*, 525 genes). Comparison of these sequences revealed a conserved core of about 250 essential genes, much smaller than either genome. In 1999, we introduced the method of global transposon mutagenesis and experimentally demonstrated that *M. genitalium* contains many genes that are nonessential for growth in the laboratory, even though it has the smallest genome known for an autonomously replicating cell found in nature. This implied that it should be possible to produce a minimal cell that is simpler than any natural one. Whole genomes can now be built from chemically synthesized oligonucleotides and brought to life by installation into a receptive cellular environment. We have applied whole-genome design and synthesis to the problem of minimizing a cellular genome.

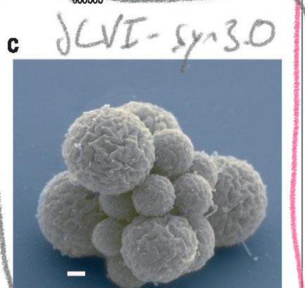
RATIONALE: Since the first genome sequences, there has been much work in many bacterial models to identify nonessential genes and define core sets of conserved genetic functions, using the methods of comparative genomics. Often, more than one gene product can perform a particular essential function. In such cases, neither gene will be essential, and neither will necessarily be conserved. Consequently, these approaches cannot, by themselves, identify a set of genes that is sufficient to constitute a viable genome. We set out to define a minimal cellular genome experimentally by designing and building one, then testing it for viability. Our goal is a cell so simple that we can determine the molecular and biological function of every gene.

RESULTS: Whole-genome design and synthesis were used to minimize the 1079–kilobase pair (kbp) synthetic genome of *M. mycoides* JCVI-syn1.0. An initial design, based on collective knowledge of molecular biology in combination with limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes that are needed for robust growth, explaining the failure of our initial design. Three more cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531 kbp, 473 genes). Its genome is smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 has a doubling time of ~180 min, produces colonies that are morphologically similar to those of JCVI-syn1.0, and appears to be polymorphic when examined microscopically.

CONCLUSION: The minimal cell concept appears simple at first glance but becomes more complex upon close inspection. In addition to essential and nonessential genes, there are many quasi-essential genes, which are not absolutely critical for viability but are nevertheless required for robust growth. Consequently, during the process of genome minimization, there is a trade-off between genome size and growth rate. JCVI-syn3.0 is a working approximation of a minimal cellular genome, a compromise between small genome size and a workable growth rate for an experimental organism. It retains almost all the genes that are involved in the synthesis and processing of macromolecules. Unexpectedly, it also contains 149 genes with unknown biological functions, suggesting the presence of undiscovered functions that are essential for life. JCVI-syn3.0 is a versatile platform for investigating the core functions of life and for exploring whole-genome design. ■

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.aad6253



Four design-build-test cycles produced JCVI-syn3.0. (**A**) The cycle for genome design, building by means of synthesis and cloning in yeast, and testing for viability by means of genome transplantation. After each cycle, gene essentiality is reevaluated by global transposon mutagenesis. (**B**) Comparison of JCVI-syn1.0 (outer blue circle) with JCVI-syn3.0 (inner red circle), showing the division of each into eight segments. The red bars inside the outer circle indicate regions that are retained in JCVI-syn3.0. (**C**) A cluster of JCVI-syn3.0 cells, showing spherical structures of varying sizes (scale bar, 200 nm).

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: chutchis@jcvi.org (C.A.H.); jcventer@jcvi.org (J.C.V.)
†These authors contributed equally to this work.
Cite this article as C. A. Hutchison III *et al.*, *Science* 351, aad6253 (2016). DOI: 10.1126/science.aad6253

Downloaded from http://science.sciencemag.org/ on June 7, 2019

NATURE 509 (pp 385-388), 2014

LETTER

doi:10.1038/nature13314

A semi-synthetic organism with an expanded genetic alphabet

Not only A-T-C-G but other two bases

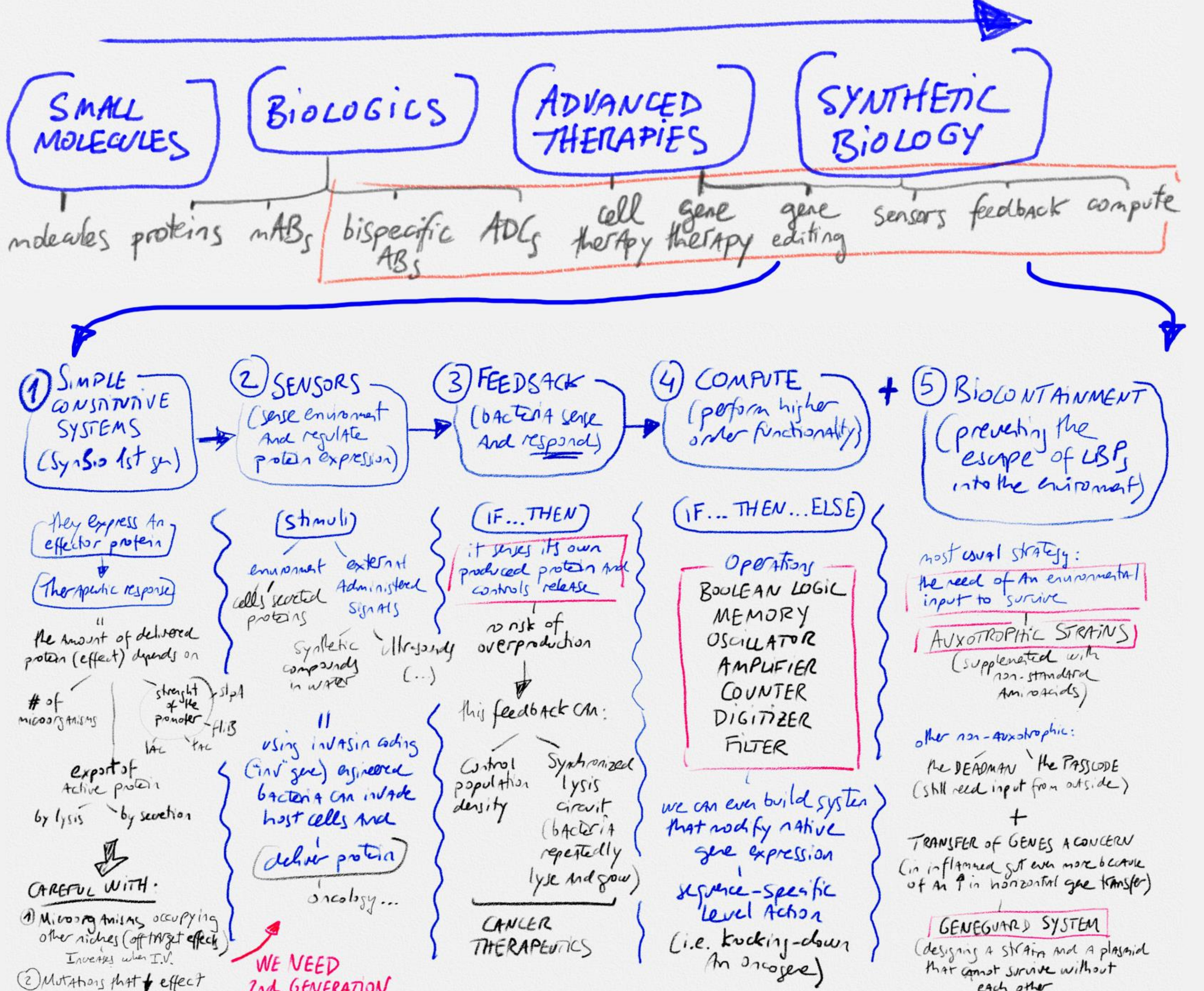
Denis A. Malyshev[1], Kirandeep Dhami[1], Thomas Lavergne[1], Tingjian Chen[1], Nan Dai[2], Jeremy M. Foster[2], Ivan R. Corrêa Jr[2] & Floyd E. Romesberg[1]

A NEW FORM OF LIFE!

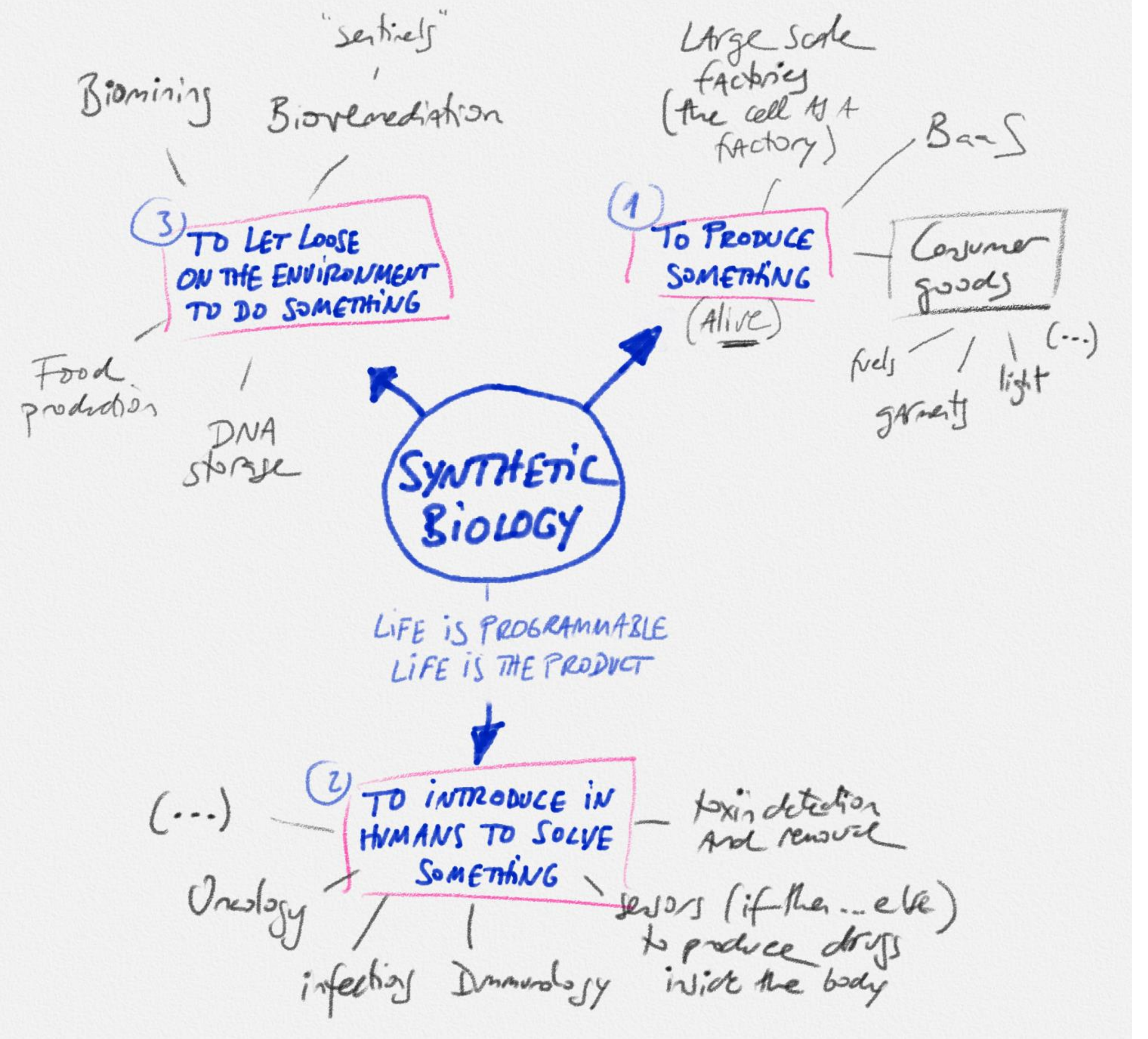
A later paper shows it can divide itself and reproduce

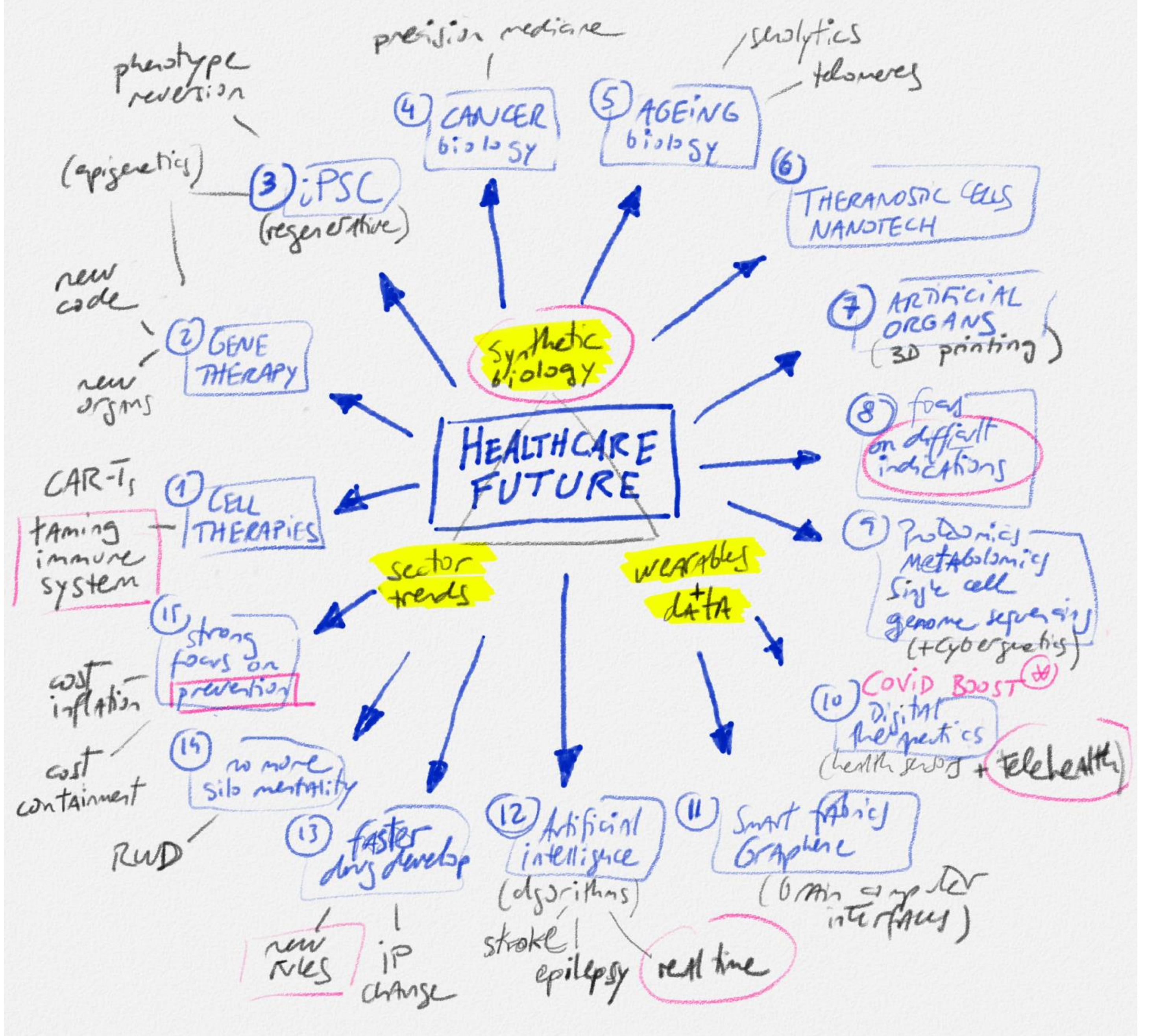
Organisms are defined by the information encoded in their genomes, and since the origin of life this information has been encoded using a two-base-pair genetic alphabet (A–T and G–C). *In vitro*, the alphabet has been expanded to include several unnatural base pairs (UBPs)[1–3]. We have developed a class of UBPs formed between nucleotides bearing hydrophobic nucleobases, exemplified by the pair formed between d5SICS and dNaM (d5SICS–dNaM), which is efficiently PCR-amplified[1] and transcribed[4,5] *in vitro*, and whose unique mechanism of replication has been characterized[6,7]. However, expansion of an organism's genetic alphabet presents new and unprecedented challenges: the unnatural nucleoside triphosphates must be available inside the cell; endogenous polymerases must be able to use the unnatural triphosphates to faithfully replicate DNA containing the UBP within the complex cellular milieu; and finally, the UBP must be stable in the presence of pathways that maintain the integrity of DNA. Here we show that an exogenously expressed algal nucleotide triphosphate transporter efficiently imports the triphosphates of both d5SICS and dNaM (d5SICSTP and dNaMTP) into *Escherichia coli*, and that the endogenous replication machinery uses them to accurately replicate a plasmid containing d5SICS–dNaM. Neither the presence of the unnatural triphosphates nor the replication of the UBP introduces a notable growth burden. Lastly, we find that the UBP is not efficiently excised by DNA repair pathways. Thus, the resulting bacterium is the first organism to propagate stably an expanded genetic alphabet.

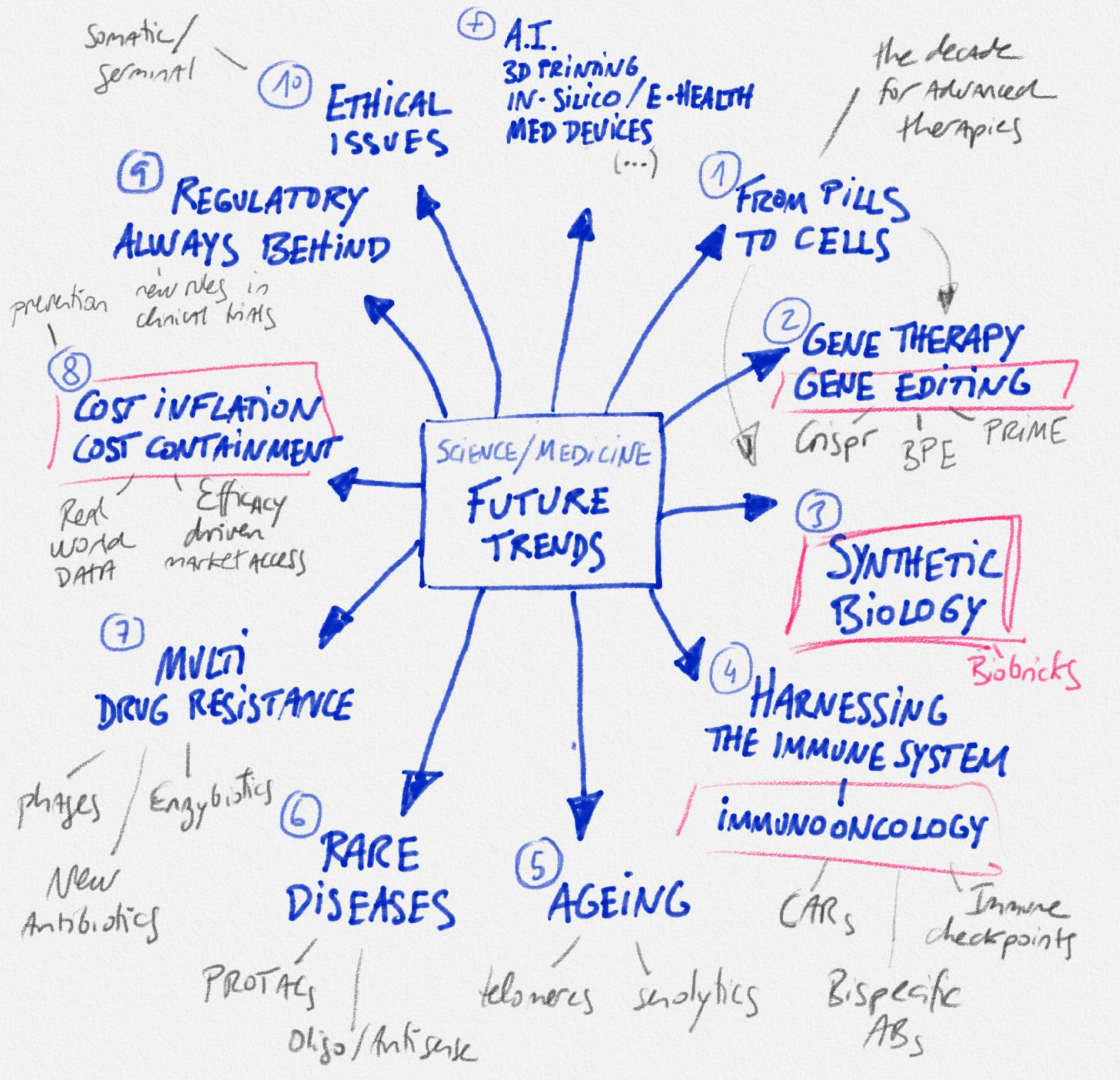
suggest that decomposition is mediated by phosphatases. As no degradation was observed upon incubation in spent media, decomposition seems to occur within the periplasm. No increase in stability was observed in cultures of single-gene-deletion mutants of *E. coli* BW25113 lacking a specific periplasmic phosphatase[19] (as identified by the presence of a Sec-type amino-terminal leader sequence), including *phoA*, *ushA*, *appA*, *aphA*, *yjjX*, *surE*, *yfbR*, *yjjG*, *yfaO*, *mutT*, *nagD*, *yggV*, *yrfG* or *ymfB*, suggesting that decomposition results from the activity of multiple phosphatases. However, the extracellular stability of [α-^{32}P]-dATP was significantly greater when 50 mM potassium phosphate (KPi) was added to the growth medium (Extended Data Fig. 3). Thus, we measured [α-^{32}P]-dATP uptake from media containing 50 mM KPi after induction of the transporter with isopropyl-β-D-thiogalactoside (IPTG) (Extended Data Fig. 4). Although induction with 1 mM IPTG resulted in slower growth, consistent with the previously reported toxicity of NTTs[17], it also resulted in maximal [α-^{32}P]-dATP uptake. Thus, after addition of 1 mM IPTG, we analysed the extracellular and intracellular stability of [α-^{32}P]-dATP as a function of time (Extended Data Fig. 5). Cells expressing *Pt*NTT2 were found to have the highest levels of intracellular [α-^{32}P]-dATP, and although both extra- and intracellular dephosphorylation was still observed, the ratio of triphosphate to dephosphorylation products inside the cell remained roughly constant, indicating that the extracellular concentrations and *Pt*NTT2-mediated influx are sufficient to compensate for intracellular decomposition.



SMALL MOLECULES
BIOLOGICS
ADVANCED THERAPIES
SYNTHETIC BIOLOGY
molecules
proteins
mABs
bispecific ABs
ADCs
cell therapy
gene therapy
gene editing
sensors
feedback
compute
1 SIMPLE CONSTITUTIVE SYSTEMS (SynBio 1st gen)
2 SENSORS (sense environment and regulate protein expression)
3 FEEDBACK (bacteria sense and respond)
4 COMPUTE (perform higher order functionality)
+
5 BIOCONTAINMENT (preventing the escape of LBPs into the environment)
they express an effector protein
Therapeutic response
the amount of delivered protein (effect) depends on
# of microorganisms
strength of the promoter
export of active protein
by lysis
by secretion
CAREFUL WITH:
1 Microorganisms occupying other niches (off target effect)
Increases when I.V.
2 Mutations that ↓ effect
stimuli
environment
external administered signals
cells' secreted proteins
synthetic compounds in water
ultrasounds
(...)
using invasin coding ("inv" gene) engineered bacteria can invade host cells and
deliver protein
oncology...
WE NEED 2nd GENERATION
IF... THEN
it senses its own produced protein and controls release
no risk of overproduction
this feedback can:
Control population density
Synchronized lysis circuit (bacteria repeatedly lyse and grow)
CANCER THERAPEUTICS
IF... THEN... ELSE
Operations
BOOLEAN LOGIC
MEMORY
OSCILLATOR
AMPLIFIER
COUNTER
DIGITIZER
FILTER
we can even build systems that modify native gene expression
sequence-specific level action
(i.e. knocking-down an oncogene)
most usual strategy:
the need of an environmental input to survive
AUXOTROPHIC STRAINS
(supplemented with non-standard aminoacids)
other non-auxotrophic:
the DEADMAN
the PASSCODE
(still need input from outside)
+
TRANSFER OF GENES A CONCERN
(in inflamed gut even more because of an ↑ in horizontal gene transfer)
GENEGUARD SYSTEM
(designing a strain and a plasmid that cannot survive without each other)


SYNTHETIC BIOLOGY
LIFE IS PROGRAMMABLE
LIFE IS THE PRODUCT
1 TO PRODUCE SOMETHING
(Alive)
Large scale factories (the cell as a factory)
BaaS
Consumer goods
fuels
garments
light
(...)
2 TO INTRODUCE IN HUMANS TO SOLVE SOMETHING
(...)
toxin detection and removal
Oncology
infections
Immunology
sensors (if-then... else)
to produce drugs inside the body
3 TO LET LOOSE ON THE ENVIRONMENT TO DO SOMETHING
Biomining
"sentinels"
Bioremediation
Food production
DNA storage


HEALTHCARE FUTURE
synthetic biology
sector trends
wearables + data
1 CELL THERAPIES
CAR-Ts
taming immune system
2 GENE THERAPY
new code
new organs
3 iPSC (regenerative)
(epigenetics)
phenotype reversion
4 CANCER biology
precision medicine
5 AGEING biology
senolytics
telomeres
6 THERANOSTIC CELLS NANOTECH
7 ARTIFICIAL ORGANS (3D printing)
8 focus on difficult indications
9 proteomics metabolomics single cell genome sequencing (+cybergenetics)
10 COVID BOOST Digital therapeutics
(health sensors + telehealth)
11 Smart fabrics Graphene
(brain computer interfaces)
12 Artificial intelligence
(algorithms)
stroke
epilepsy
real time
13 faster drug develop
new rules
IP change
14 no more silo mentality
cost containment
RWD
15 strong focus on prevention
cost inflation



SCIENCE/MEDICINE
FUTURE TRENDS
1 FROM PILLS TO CELLS
the decade for advanced therapies
2 GENE THERAPY GENE EDITING
Crispr
BPE
PRIME
3 SYNTHETIC BIOLOGY
Biobricks
4 HARNESSING THE IMMUNE SYSTEM
IMMUNOONCOLOGY
CARs
Bispecific ABs
Immune checkpoints
5 AGEING
telomeres
senolytics
6 RARE DISEASES
PROTACs
Oligo/Antisense
7 MULTI DRUG RESISTANCE
phages
Enzybiotics
New Antibiotics
8 COST INFLATION COST CONTAINMENT
prevention
Real World DATA
Efficacy driven market access
9 REGULATORY ALWAYS BEHIND
new rules in clinical trials
10 ETHICAL ISSUES
Somatic/germinal
+ A.I.
3D PRINTING
IN-SILICO/E-HEALTH
MED DEVICES
(...)

ARTICLE

https://doi.org/10.1038/s41467-020-18008-4 OPEN

A machine learning Automated Recommendation Tool for synthetic biology

Tijana Radivojević[1,2,3], Zak Costello[1,2,3], Kenneth Workman[1,3,4] & Hector Garcia Martin[1,2,3,5]✉

Synthetic biology allows us to bioengineer cells to synthesize novel valuable molecules such as renewable biofuels or anticancer drugs. However, traditional synthetic biology approaches involve ad-hoc engineering practices, which lead to long development times. Here, we present the Automated Recommendation Tool (ART), a tool that leverages machine learning and probabilistic modeling techniques to guide synthetic biology in a systematic fashion, without the need for a full mechanistic understanding of the biological system. Using sampling-based

ACS Synthetic Biology

Viewpoint

Cite This: ACS Synth. Biol. 2019, 8, 1474–1477

pubs.acs.org/synthbio

Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation

Pablo Carbonell,[†] Tijana Radivojevic,[‡,∥] and Héctor García Martín[*,‡,§,∥,⊥]

[†]Manchester Synthetic Biology Research Centre for Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology and School of Chemistry, University of Manchester, Manchester M1 7DN, United Kingdom

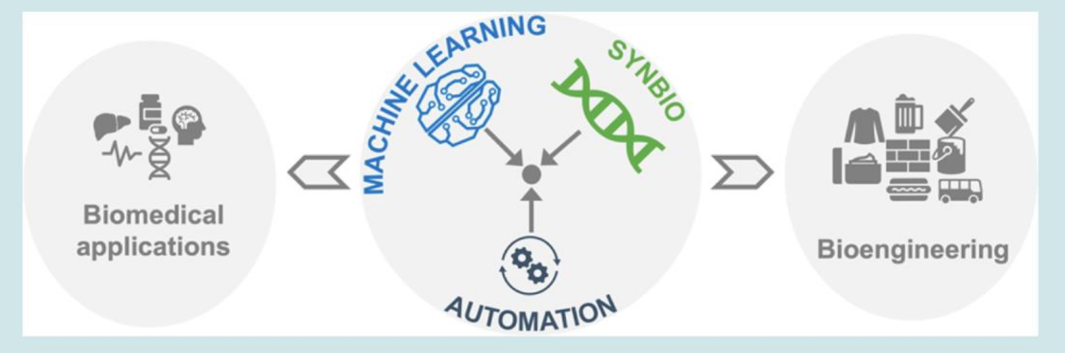
[‡]DOE Agile BioFoundry, Emeryville, California 94608, United States

[§]DOE Joint BioEnergy Institute, Emeryville, California 94608, United States

[∥]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

[⊥]BCAM, Basque Center for Applied Mathematics, 48009 Bilbao, Spain

ABSTRACT: Our inability to predict the behavior of biological systems severely hampers progress in bioengineering and biomedical applications. We cannot predict the effect of genotype changes on phenotype, nor extrapolate the large-scale behavior from small-scale experiments. Machine learning techniques recently reached a new level of maturity, and are capable of providing the needed predictive power without a detailed mechanistic understanding. However, they require



large amounts of data to be trained. The amount and quality of data required can only be produced through a combination of