Reading & Writing Omics
5-Dec-2019, 12-1 pm, George Church
Duke Applied Genomics & Precision Medicine Forum

Full COI list: v.ht/PHNc
Reading & Writing Omes
(from the viewpoint of prevention & therapy)

Relevant commercial interests:
3rd Gen Seq: ONT, Roche, Roswell
2nd Gen Seq: Nebula Genomics, Veritas Genetics, Illumina
In situ: Readcoor, Bruker
Edit: Editas Medicine, IntelliaTx, CrisprTx, Egenesis
Tissues in vitro: Cellino, GCTx
Aging & complex diseases: Rejuvenatebio

(full COI see: v.ht/PHNc)
World’s only fully open access human Genomic, Environmental, Trait data & cells

NIST + FDA + PGP = Genomeinabottle.org
Standards: Ashkenazi & Han trios

NIH ENCyclopedia Of DNA Elements
11 isogenic cell types.

Critical Assessment of Genome Interpretation

US, Canada, UK, Austria, Korea, China, India
<table>
<thead>
<tr>
<th>Therapy</th>
<th>Source</th>
<th>Genes</th>
<th>Disease</th>
<th>US/yr</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxturna</td>
<td>Spark</td>
<td>AAV2/RPE65</td>
<td>LCA-blindness</td>
<td>80</td>
<td>425k/eye</td>
</tr>
<tr>
<td>Imlygic</td>
<td>Amgen</td>
<td>HSV1/GM-CSF</td>
<td>Melanoma</td>
<td>7200</td>
<td>780k/6mo</td>
</tr>
<tr>
<td>Kymriah</td>
<td>Novartis</td>
<td>CAR(to CD19)</td>
<td>Leukemia</td>
<td>9600</td>
<td>475k</td>
</tr>
<tr>
<td>Yescarta</td>
<td>Kite Pharma</td>
<td>CAR(to CD19)</td>
<td>Lymphoma</td>
<td>8000</td>
<td>373k</td>
</tr>
<tr>
<td>Zolgensma</td>
<td>Novartis</td>
<td>AAV9/SMN1</td>
<td>SMA</td>
<td>500</td>
<td>2.1M</td>
</tr>
<tr>
<td>Spinraza</td>
<td>Biogen</td>
<td>ASO</td>
<td>SMA</td>
<td>500</td>
<td>4M/10yr</td>
</tr>
<tr>
<td>Multiplex</td>
<td>Rejuvenate</td>
<td>AAV/3genes</td>
<td>Multiplex</td>
<td>2,800,000</td>
<td>2k ?</td>
</tr>
</tbody>
</table>

1) Common diseases: (age-affected) spread R&D costs
2) Rare diseases: Prevention via dating/matchmaking
Nucleic Acid Read / Write improving exponentially 2-10-fold better than Moore’s law

- 1E-3 to 1E+7
- 1980 to 2015
- Oligo bp/$
- Seq bp/$

1984 HGP haploid 1E-2 error rate

Multiplex NGS in 2003

2009-12 diploid PGP 1E-7 error rate

Multiplex Chip-Syn 2004

$0
Molecular & cell multiplexing
1E2 to 1E12 reactions per droplet
(not automation or parallelism)

1987 US #4,942,124 Multiplex sequencing.
2009 Multiplex Automated Genome Engineering. Nature
2018 Chan et al. Multiplexed testing of pooled donor cells using SNPs and NGS. Genome Medicine

![Diagram of multiplexing and de-multiplexing](image)
Addition, subtraction, precise-editing, epigenetics

- Nucleases (error-prone)
- De-aminases $C \rightarrow T$, $A \rightarrow G$
- Recombinases & Primer-Editor (precise & flexible)
Recombinases & Integrases

- Red recombinase: Gabe Filsinger & Tim Wannier found interaction with Single-strand binding proteins (SSB). No custom protein or RNA (just donor DNA)

- Small S/Y integrases: Ben Weinberg, ML-Syn. Custom protein (no RNA)
GP-Write: precise hyperediting (MAGE)

4 Million bp Genome Recoding: 63 & 57 codon types

1. Non-standard amino acids (NSAA)

2. Genetic & Metabolic Isolation

3. Multi-Virus resistance

Science 2016 Ostrov, et al
Nature 2015 Mandell et al
Science 2013 Lajoie et al
### Genomically Recoded Organisms (GRO)

<table>
<thead>
<tr>
<th>5' base</th>
<th>2nd base</th>
<th>3' base</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>F</td>
<td>22445</td>
<td>16567</td>
</tr>
<tr>
<td>L</td>
<td>18824</td>
<td>12064</td>
</tr>
<tr>
<td>L</td>
<td>18488</td>
<td>10 RF1</td>
</tr>
<tr>
<td>S</td>
<td>9620</td>
<td>1 RF2</td>
</tr>
<tr>
<td>U</td>
<td>30204</td>
<td>6948</td>
</tr>
<tr>
<td>C</td>
<td>11672</td>
<td>21888</td>
</tr>
<tr>
<td>P</td>
<td>11396</td>
<td>28424</td>
</tr>
<tr>
<td>L</td>
<td>9476</td>
<td>2737</td>
</tr>
<tr>
<td>I</td>
<td>7401</td>
<td>7273</td>
</tr>
<tr>
<td>T</td>
<td>13140</td>
<td>4744</td>
</tr>
<tr>
<td>V</td>
<td>29866</td>
<td>4 RF2</td>
</tr>
<tr>
<td>G</td>
<td>14953</td>
<td>1232</td>
</tr>
<tr>
<td>A</td>
<td>71864</td>
<td>20688</td>
</tr>
<tr>
<td>C</td>
<td>11425</td>
<td>28424</td>
</tr>
<tr>
<td>N</td>
<td>17534</td>
<td>28424</td>
</tr>
<tr>
<td>K</td>
<td>20842</td>
<td>28424</td>
</tr>
<tr>
<td>R</td>
<td>39234</td>
<td>28424</td>
</tr>
<tr>
<td>S</td>
<td>7273</td>
<td>28424</td>
</tr>
<tr>
<td>R</td>
<td>1491</td>
<td>28424</td>
</tr>
<tr>
<td>T</td>
<td>19572</td>
<td>28424</td>
</tr>
<tr>
<td>V</td>
<td>13937</td>
<td>28424</td>
</tr>
<tr>
<td>A</td>
<td>31603</td>
<td>28424</td>
</tr>
<tr>
<td>C</td>
<td>39234</td>
<td>28424</td>
</tr>
<tr>
<td>G</td>
<td>45816</td>
<td>28424</td>
</tr>
</tbody>
</table>

- **2 +1 codons**
  (Fredens..Chin 2019)
- **6 +1 codons**
  4 Mbp synthesized
  3kb → 50 kb
- **Ostrov, Norville, Guell, .. Church 2016**
<table>
<thead>
<tr>
<th>Ref.</th>
<th>Disease</th>
<th>Clinical Onset</th>
<th>Organoid</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 Cardiomyopathy</td>
<td></td>
<td>1y</td>
<td>4 wk</td>
</tr>
<tr>
<td>2018 Late onset Alzheimer’s</td>
<td></td>
<td>70y</td>
<td>2 wk</td>
</tr>
<tr>
<td>2019 Schizophrenia</td>
<td></td>
<td>20y</td>
<td>4 wk</td>
</tr>
<tr>
<td>2019 Bipolar</td>
<td></td>
<td>20y</td>
<td>4 wk</td>
</tr>
</tbody>
</table>

The most common disease (affects rare & complex diseases) → 9 Pathways

- Clearance of senescent cells
- Caloric restriction
- Reactivating protein regulation
- Epigenetic drugs
- Restoring mito function
- Stem cell addition
- Inflammation reduction, blood rejuvenation
- Reducing cells with genomic damage
- Telomere reactivation

Non-cell autonomous aging reversal gene therapies

GenAge
The Ageing Gene Database
305 Human, 126 Mouse
838 Worm, 170 Fly
João Pedro de Magalhães

FGF21: glucose handling,
sTGFβR2: Soluble form
αKlotho: intracellular calcium, serum mineral-ion homeostasis

→ 5 aging diseases via combinatorial AAV gene therapy

1. High fat diet obesity
2. Type 2 Diabetes
3. Osteoarthritis
4. Cardiac damage recovery
5. Kidney disease

Noah Davidsohn et al. PNAS Nov-2019
Protein Design (AAV & Ab): Multiplex deep sequence testing

DNA Synthesis → Machine Learning → Selection → DNA sequencing

- Capsid packaging
- Immune evasion
- In vivo delivery
- Tissue specificity

Pierce Ogden
Eric Kelsic
Science Nov 2019
## Mapping the space of tissue tropism modifications

<table>
<thead>
<tr>
<th></th>
<th>Blood</th>
<th>Heart</th>
<th>Liver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kidney</td>
<td><img src="image" alt="Heatmap" /></td>
<td><img src="image" alt="Heatmap" /></td>
<td><img src="image" alt="Heatmap" /></td>
</tr>
<tr>
<td>Spleen</td>
<td><img src="image" alt="Heatmap" /></td>
<td><img src="image" alt="Heatmap" /></td>
<td><img src="image" alt="Heatmap" /></td>
</tr>
<tr>
<td>Lung</td>
<td><img src="image" alt="Heatmap" /></td>
<td><img src="image" alt="Heatmap" /></td>
<td><img src="image" alt="Heatmap" /></td>
</tr>
</tbody>
</table>

Position

570

Position

605
VUS-pipeline: Full Transcription Factor Library

Collaborators

1,576 genes → 1,748 human TF ORFs.
Gateway-compatible vectors (+1200 Sheynkman)

1. Many cell types: Screen for loss of stem cell identity, TRA 1-60

2. Make specific cell types via transcriptome targeting

Ng, Khoshakhlagh, Shipman, Swiersy, Appleton, Huang, Saylor, Trono, Taipale, Hill, Vidal, Busskamp, Church (2017) in prep
Oligodendrocyte
Myelin wrapping Axons
Human organoids in mice

We observed enhanced motor function after OPC injection

Healthy mouse

Shiverer: OPC Injection

Shiverer: Mock Injection

Hind limbs far apart: Score 4

Hind limbs far apart: Score 4

Hind limbs close together: Score 3
Pig organs in primates. Multiplex (43-plex) germline editing

**Sugars:** GGTA1, CMAH, β4GalNT2

**Clotting:** human TFPI, TBM, EPCR, vWF

**MHC:** HLA/SLA class I, II

**Immune functions:** human CTLA4-Ig, HLA-E/G/Cw3 inhibit NK cells

**Complement:** regulatory genes CD46, CD55, CD59

**Porcine ERVs:** 25 in normal fibroblasts

**MGH:**
Jim Markmann

Luhan Yang, et al 2015 Science
Dong Niu, et al. 2017 Science
Genome engineering for enhanced organs

Resistance to: pathogens, senescence, cancer, immunity, cryopreservation, DNA-damage

*Lithobates sylvaticus* N.Amer. wood frog -14°C
*Spermophilus parryii*, Arctic squirrel -3°C
*Gynaephora groenlandica*, Arctic moth -60°C
*Polypedilum vanderplanki*, arid midge 3°K
*Tardigrada*, broadly distributed 3°K

Glucose + urea
Trehalose, Late Embryo Abundant, anti-oxidant, heat-shock proteins, AFPs, PUFAs, polyamines, etc.
DNA-nick toxicity: bFGF & PTFα preserves highly modified cells

Niu et al. 2017 Science
## Multiplex Editing Genome-wide

<table>
<thead>
<tr>
<th>Category</th>
<th>#/ hum cell</th>
<th>Length</th>
<th>Total bp</th>
<th>Previous mutation attempts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERV</td>
<td>25</td>
<td>7500</td>
<td>2.4E+7</td>
<td>Cas9 Yang .. Science 2015, 2017</td>
</tr>
<tr>
<td>Telomeres</td>
<td>46</td>
<td>8000</td>
<td>3.7E+5</td>
<td>TERT Ramunas .. FASEB J 2015</td>
</tr>
<tr>
<td>rDNA repeats</td>
<td>300</td>
<td>43000</td>
<td>1.3E+7</td>
<td>I-CreI Paredes .. Genetics 2009</td>
</tr>
<tr>
<td>UCE</td>
<td>855</td>
<td>200</td>
<td>1.7E+5</td>
<td>Cas9 Schneider, Biorxiv 2019</td>
</tr>
<tr>
<td>LINEs</td>
<td>26,000</td>
<td>7000</td>
<td>1.8E+8</td>
<td>dABE Smith .. Biorxiv 2019</td>
</tr>
<tr>
<td>Centromeres</td>
<td>1.0E+6</td>
<td>171</td>
<td>1.7E+8</td>
<td>Cas9 Adikusuma .. Mol Ther 2017</td>
</tr>
<tr>
<td>SINEs (Alu)</td>
<td>1.5E+6</td>
<td>280</td>
<td>4.2E+8</td>
<td>Natural variants</td>
</tr>
<tr>
<td>SSR</td>
<td>3.0E+6</td>
<td>16</td>
<td>4.8E+7</td>
<td>Cas9 [14] Monteys .. Mol Ther 2017</td>
</tr>
<tr>
<td>Triplex sites</td>
<td>1.7E+7</td>
<td>20</td>
<td>3.4E+8</td>
<td></td>
</tr>
</tbody>
</table>

### Repeats involved in:

Wang et al. Cell Cycle 2011  **cancer (Cen, Alu)**
DeCecco et al. Nature 2019  **senescence, inflammation (L1)**
Singer et al. Trends Neurosci. 2010  **neurogenesis (L1)**
Enabling large-scale genome editing by reducing DNA nicking & off-target

Deamination

\[ 5' - ATTCTACCAGAGGTACAAGGAGG - 3' \]

\[ 5' - ATT \underline{T} TATTTAGAGGTACAAGGAGG - 3' \]

\[ 5' - ATTCTGCCAGAGGTACAAGGAGG - 3' \]

1. No nick, dCas9
2. No MMR
3. No U-glycosylase
4. Anti-apoptotic molecules
5. Growth factors (bFGF)

Nick-less deaminase targeting of LINE-1 in human 293T cells (PGP1 iPSC)

- **dABE**: 50.61% x 26100 LINE-1 = 13,470 edits (27,000 diploid)
- **dCBE**: 22% x 26100 LINE-1 = 5742 edits
Storing data in DNA

<455
ExaBytes/g

Church, Gao, Kosuri 2012
Science

Activity, Behavior, Development
12 billion cells * 40 mutant gRNAs → 4 trillion bits
Recording in DNA *in vivo*: Developmental lineage

4 trillion bits stored

Kalhor, Mali, Church
2016 Nature Methods
2018 Science

gRNAs
in situ
### Multiplex cell analysis instruments

<table>
<thead>
<tr>
<th>nm</th>
<th>Plex</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>90,000</td>
<td>&gt;1000 NA</td>
<td>Spatial Transcriptomics</td>
</tr>
<tr>
<td>5,000</td>
<td>40 Ab</td>
<td>CyTOF (flow)</td>
</tr>
<tr>
<td>250</td>
<td>50 Ab</td>
<td>CyCIF (in situ)</td>
</tr>
<tr>
<td>250</td>
<td>&gt;1000 NA+Ab</td>
<td>ReadCoor (in situ) targeted or un</td>
</tr>
<tr>
<td>12</td>
<td>&gt;1000 NA+Ab</td>
<td>Bruker Vutara (in situ) targeted or un</td>
</tr>
</tbody>
</table>
Normal, synthetic, pathological tissues
in situ
DNA, RNA, protein
Fluor microscopy
(hyb, lig, pol)

ReadCoor (Church lab)
&
Bruker Vutara (Wu lab)
12 nm precision
biplane & lightsheet

Vutara
$400k

1000-fold thicker samples than

Novaseq
$985k
Subcellular & multicellular
*in situ* RNA sequencing  PGP1 fibroblasts

Lee, Daugharthy, .. Church
Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLoS Gen 2018

Guy Nir
Huy Nguyen
Shyamtanu Chattoraj
Brian Beliveau

Ting Wu, Peng Yin, Erez Lieberman-Aiden, Marc Marti-Renom
Chr. 19 step sizes: 1300 to 400 kb
Chr. 5: all steps are 250 kb
Chr. 3: all steps are 500 kb
OligoSTORM

Nir, Farabella, Perez Estrada, Ebeling ... Wu 2018
OligoSTORM

Nir, Farabella, Perez-Estrada, Ebeling ... Wu 2018
Male Lymphocytes

600 kb-1Mb/target (876 kb average)
5,000 Oligopaints oligos/target
7-70 Mb between targets

Huy Nguyen
Shyamtanu Chattoraj
David Castillo/ Marc Marti-Renom
Ting Wu
Oligonucleotide-labeled Protein Recognition

DNA Exchange Imaging (DEI): Super-resolution & Multiplex ML-Syn of nanobodies

hybribody.com/contenu/synthetic-vhh-library-menu/take-advantage-of-vhh-antibody-properties
Proteins: *in situ* multiplexing with DNA-Exchange

Binder library & assay design

2 different Target Libraries
- Cytokines
- Transcription Factors

Target Library | Barcode | Binder Library

Target – Binder Interaction

Reporter

NGS barcodes to see what interacts

6 different Binder Libraries
- Adnectins
- Affibodies
- Affimers
- Alphabodies
- Fynomers
- Nanobodies
Machine learning on many Binder-Target interactions

- Adnectins
- Affibodies
- Affimers
- Alphabodies
- Fynomers
- Nanobodies

Gleb Kuznetsov & Pierce Ogden
PGP collaborators
NIH NHGRI CEGS
Center for Genomically Engineered Organs

Boyden, Church, Lewis, Wu

DARPA, NSF, DOE GP-Write

Isaacs, Wang, Silver, Way, Boeke

NIH Transformative Awards

Molecular Recording
Boyden, Kording

Human iPSC: Alzheimer’s Schizophrenia & Bipolar
Yankner

arep.med.harvard.edu/gmc/ELSI_gc.html pged.org
Mammalian Genome Recoding Team

Eriona Hysoli
Yuting Chen

Harvard Undergrads

QFASTR
SIAT
Richie Kohman

Stephen Casper
Kevin Yang
Germline vs Somatic Gene Therapy

1. Billion-fold lower off-target (due to cell #)
2. Clonality allows checking (additional million-fold)
3. Immune Tolerance
4. 100% Delivery to every cell
5. $0 to subsequent generations

Disadvantages:
1. Very long term clinical trials >20 years
2. 130 newborns/yr vs 7.7 billion total
Non-functional UAG thwarts 4 / 6 Viruses

Ma & Isaacs (2016)
Cell Syst. 3: 199–207.
Carbon sequestration (not just slowing losses)

Human use: 10 Gton/yr  Tundra soil carbon: 1,400 Gton
Atmosphere: 850 Gton  Tropical forests: 375 Gton

1. Alternatives to meat
2. Virus-resistant algae, segregate CHO from Fe/P
3. Keystone species to switch back trees to grass
   increase albedo, photosynthesis, winter snow pack

2008  |  Nicholas Wade  |  The New York Times
2012  |  REGENESIS       |  "George Church is one of the most influential scientists in the world."
            |  GEORGE CHURCH AND ED REGIS
2013  |  NATIONAL GEOGRAPHIC |  NATIONAL GEOGRAPHIC
              |  REVIVING EXTINCTION SPECIES
4-Jul-2017  |  WOOLLY         |  "The True Story of the Quest to Revive One of History's Most Iconic Extinct Creatures"
            |  BEN MEZRICHT
USDA per 100 grams
**Bison** = 109 calories, 1.8 grams fat
**Beef** = 291 calories, 24 grams fat

medium.com/@eriona.hysolli/an-american-russian-collaboration-to-repopulate-siberia-with-woolly-mammoths-or-something-similar-9cbac4e985cb
Targeted Cytosine Deaminases


2016 Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Komor … Liu Nature C→T Cas9-BE


2019 Cytosine base editor generates substantial off-target SNVs in mouse embryos. (20-fold APOBEC1 vs TadA) Zui … Yang. Science
Adenosine Deaminase

2017 Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Gaudelli ... Liu. Nature

![Diagram of Adenosine Deaminase and base editing process]

<table>
<thead>
<tr>
<th>ABE</th>
<th>E. coli TadA amino acid</th>
<th>TadA state</th>
<th>Linker 1 length</th>
<th>Linker 2 length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>W H P R L A D H A S D R E I K</td>
<td>Monomer</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>W H P R L A N H A S D R E I K</td>
<td>Monomer</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>W H P R L V N H A S Y R V I K</td>
<td>Monomer</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>W H P R L V N H A S Y R V I K</td>
<td>Monomer</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>W H P R L V N H A S Y R V I K</td>
<td>Monomer</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>W H P R L V N H A S Y R V I K</td>
<td>Monomer</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>W H P R F V N Y A S Y R V F K</td>
<td>Monomer</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>4.3</td>
<td>W H P R F V N Y N S Y R V F K</td>
<td>Homodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>5.1</td>
<td>W L P L F V N Y A C Y R V F N</td>
<td>Homodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>5.3</td>
<td>W L P L F V N Y A C Y R V F N</td>
<td>Homodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>6.3</td>
<td>W L S L F V N Y A C Y R V F N</td>
<td>Homodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>6.4</td>
<td>W L S L F V N Y A C Y R V F N</td>
<td>Homodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>7.4</td>
<td>R L A L F V N Y A C Y R V F N</td>
<td>Heterodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>7.8</td>
<td>L L A L F V N Y N C Y R V F N</td>
<td>Heterodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>7.9</td>
<td>L L A L F V N Y N C Y P V F N</td>
<td>Heterodimer</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>7.10</td>
<td>R L A L F V N Y A C Y P V F N</td>
<td>Heterodimer</td>
<td>32</td>
<td>32</td>
</tr>
</tbody>
</table>
Reverse Transcriptase