Concept Clearance: Technology Development for Single-Molecule Protein Sequencing

Tina Gatlin, Valentina Di Francesco, Ajay Pillai, Adam Felsenfeld
Division of Genome Sciences, NHGRI, NIH
May 18, 2020 NACHGR Council
Purpose

• Accelerate innovation and development in single-molecule protein sequencing (SMPS)
  • Achieve tech advances to the level where SMPS can be used for genome-wide surveys;
  • Improve speed, sensitivity, quantitation and accuracy to use routinely in genome biology and function
  • Apply lessons learned from DNA sequencing to proteome at scale
  • Explore feasibility within budget constraints
Background

• Human proteome is extremely complex
  • Typical cell expresses >10,000 unique proteins
  • Can contain 100X as many proteoforms
  • Dynamic range 7 to 10 orders of magnitude

• Two main approaches
  • Affinity reagents
  • Mass spectrometry (MS)

• Currently, no technologies for routine proteome-scale sequencing and quantification
SMPS – Why now?

• Recent promising technological advances
  • Nanopore, Edman chemistry; companies
• Significant opportunity to advance state-of-art
• Facilitate low abundance protein detection and single cell analysis at high throughput.
• Enable improved cataloguing of protein gene products and “missing proteins”
SMPS - Why NHGRI?

- Extension of DNA seq tech into proteome world
  - scale, towards quantitation and de novo sequencing
- Expand understanding of genome biology and function
  - Genotype to phenotypes
  - Enable single cell genomic analysis
  - Establish roles of genes in pathways and networks
  - Multi-omic molecular diagnostics
Scope and Objectives

- Support investigator-initiated novel research with aim to significantly advance SMPS technologies
  - Novel, high-risk; not incremental advances
- Example techniques appropriate for development:
  - Nanopore
  - Edman-like degradation with parallel measurements
  - Fluorescence-based measurements
  - Recognition tunneling
  - Other technologies that have potential to scale genome-wide
- Not appropriate
  - Mass spectrometry
  - Technologies that are not on a path to scale
Mechanisms and Budget

<table>
<thead>
<tr>
<th></th>
<th>FY21</th>
<th>FY22</th>
<th>FY23</th>
<th>FY24</th>
<th>FY25</th>
</tr>
</thead>
<tbody>
<tr>
<td>R01</td>
<td>2.0</td>
<td>4.0</td>
<td>6.0</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>R21</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>SBIR</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>3.5</td>
<td>7.0</td>
<td>9.0</td>
<td>6.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>

dollars in millions  Grand Total = $29M with SBIR ($21M without SBIR)

- RFA R01 (Research Project) up to $500K direct costs/year; project period up to 3 years
- RFA R21 (Exploratory/Developmental Research) up to $200K DC/year; project period up to 2 years
- RFA R43/R44 SBIR; up to total costs $250K for Phase I, $2M for Phase II
- Seek sign-on from other ICs
  - NHGRI is small player in proteomics – 1-2% of NIH
NHGRI Technology Development Program

Sept 2019 Council
R01/R21
Grow to $45M/yr

SMPS
R01/R21
Grow to $7M/yr

council approved

Council approved
$1.5 M Coordinating Center RFA
$8 M Synthetic Tech RFA set

Sequencing Tech
$7 M Approved growth
Current RFA set

Unsolicited (PA/PAR)
Genomic Tech Dev
“Parent” FOAs

thank you Mike Smith
Questions / Discussion