Identifying core competencies in research reproducibility and -omics for graduate students to have upon graduation

Reproducibility and Omics Taskforce

January 6, 2017

Contents

LaTeX

1 Research Reproducibility Across the Research Process

Key Concepts: Rigor in Design and Implementation and Transparency in Dissemination.
The following is a list of competencies for students to have or to be aware of.

0. Awareness of the current problem of irreproducible research: why most published research findings are false.

1. **Study Design & Implementation**

   (a) **RIGOR** in Design (rigorous methods, specifying research questions and hypotheses in advance, outlining analyses including variable definitions and table shells, documentation).

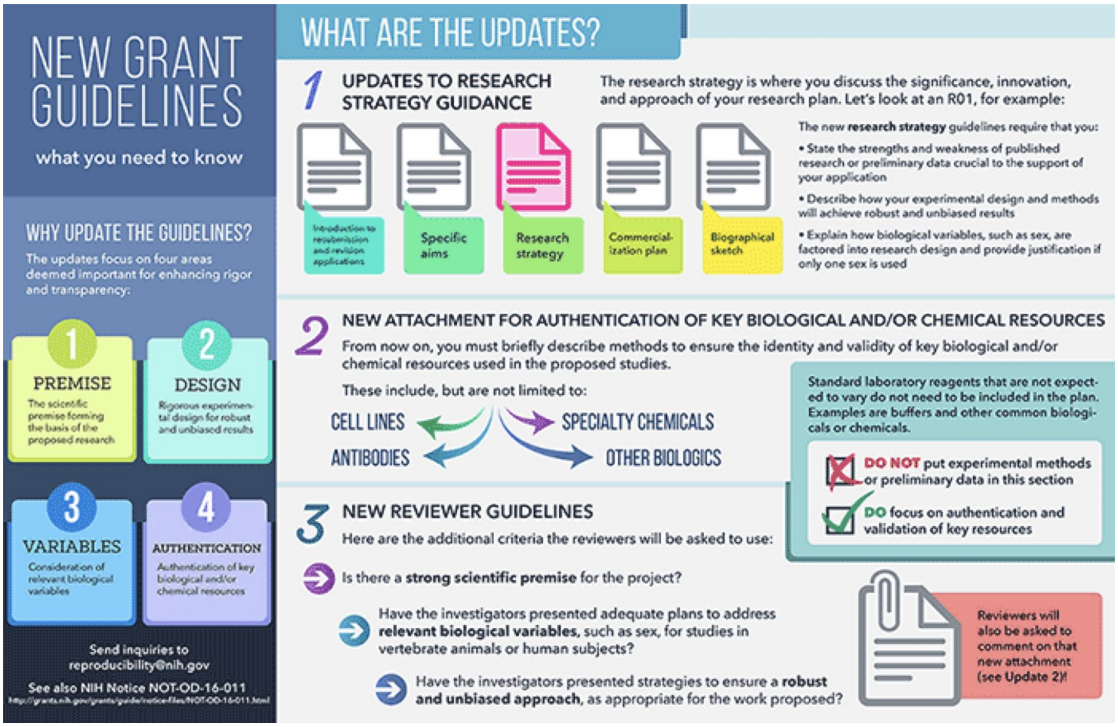
   (b) **NIH Key Areas**

      i. Strong Scientific **Premise**: Describe the general strengths and weaknesses of the prior research, consider the general strengths and weaknesses including attention to the rigor of the previous experimental designs, as well as the incorporation of relevant biological variables and authentication of key resources.

      ii. Rigorous Experimental **Design** for robust and unbiased results: The strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results. This includes full transparency in reporting experimental details so that others may reproduce and extend the findings.

      iii. Consideration of Relevant **Biological Variables** such as sex, age, weight, and underlying health conditions, are often critical factors affecting health or disease. In particular, sex is a biological variable that is frequently ignored in animal study designs and analyses, leading to an incomplete understanding of potential sex-based differences in basic biological function, disease processes and treatment response. NIH expects that sex as a biological variable will be factored into research designs, analyses, and reporting in vertebrate animal and human studies. Strong justification must be provided for applications proposing to study only one sex.

iv. **Authentication** of key biological and/or chemical resources: If using these, make sure to describe the identity and validity of key biological and/or chemical resources used in the study including cell lines, antibodies, specialty chemicals, and other biologics.



(c) Reframing the research question with **competing hypotheses**: A good approach is to consider multiple competing hypotheses (rather than solely null and alternative (Platt, 1964).

(d) **Pre-Registration**: submit research rationale, hypotheses, design, and analytic strategy to journal for peer review before beginning study.

i. Registration of clinical trials at clinicaltrials.gov.

(e) Detailed plan for data analysis completed *before* implementing study.

i. **Clarify exploratory and confirmatory analyses before implementing study.**

(f) Keep a **clear, detailed record** of methods throughout implementation to ensure transparency in dissemination.

i. Tools:

A. Create structured and shared workspace, e.g., wikis inc. OSF.

B. Support version control, e.g., GITHUB & OSF.

C. Literate programming tools, e.g., Sweave & RMarkdown.

ii. Use published checklists in study design such as STROBE, CONSORT, etc.

2. Analysis

   (a) **Document** all phases of data cleaning and data analysis to ensure replicability and computational reproducibility (*see below for definitions*).

   (b) Detailed plan for data analysis created ***before*** looking at the data

      i. Researcher degrees of freedom: All data processing and analytical choices made after seeing and interacting with your data leads to false-positive inflation.

   (c) Exploratory vs. Confirmatory approach decided ***before*** data are collected.

      i. Exploratory: Interested in exploring possible patterns/relationships in data to develop hypotheses.

      ii. Confirmatory: Have a specific hypothesis you want to test.

   (d) Understanding (awareness) of **literate programming** concept: Treat a program as a piece of literature, addressed to human beings rather than a computer.

3. Dissemination

   (a) **TRANSPARENCY** in Dissemination.

   (b) Researchers should be able to describe their methods so the reader has a clear understanding of how the data was gathered and analyzed to achieve unbiased results. Make sure manuscript includes enough detail to ensure:

      - *Empirical reproducibility*: We have enough information to rerun the experiment or survey the way it was originally conducted.
      - *Replicability*: We use original exact methods and analyses, but collect new data, and we get the same statistical results.
      - *Computational reproducibility*: If we took your data and code/analysis scripts and reran it, we can reproduce the numbers/graphs in your paper.

   (c) **Published Checklists** to ensure reproducibility (STROBE, CONSORT, etc.)

   (d) **Reporting Standards** (from NIH LINK)

      i. Encourage the use of **community-based standards** (such as nomenclature standards and reporting standards like ARRIVE), where applicable.

      ii. **Replicates**: Require that investigators report how often each experiment was performed and whether the results were substantiated by repetition under a range of conditions. Sufficient information about sample collection must be provided to distinguish between independent biological data points and technical replicates.

iii. **Statistics**: Require that statistics be fully reported in the paper, including the statistical test used, exact value of N, definition of center, dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals).

iv. **Randomization**: Require authors to state whether the samples were randomized and specify method of randomization, at a minimum for all animal experiments.

v. **Blinding**: Require authors to state whether experimenters were blind to group assignment and outcome assessment, at a minimum for all animal experiments.

vi. **Sample-size estimation**: Require authors to state whether an appropriate sample size was computed when the study was being designed and include the statistical method of computation. If no power analysis was used, include how the sample size was determined.

vii. **Inclusion and exclusion criteria**: Require authors to clearly state the criteria that were used for exclusion of any data or subjects. Include any similar experimental results that were omitted from the reporting for any reason, especially if the results do not support the main findings of the study. Describe any outcomes or conditions that were measured or used and are not reported in the results section.

(e) Data provenance & sharing: Researchers should at least be share-curious in regards to their data. (avoid[1] being a data parasite).

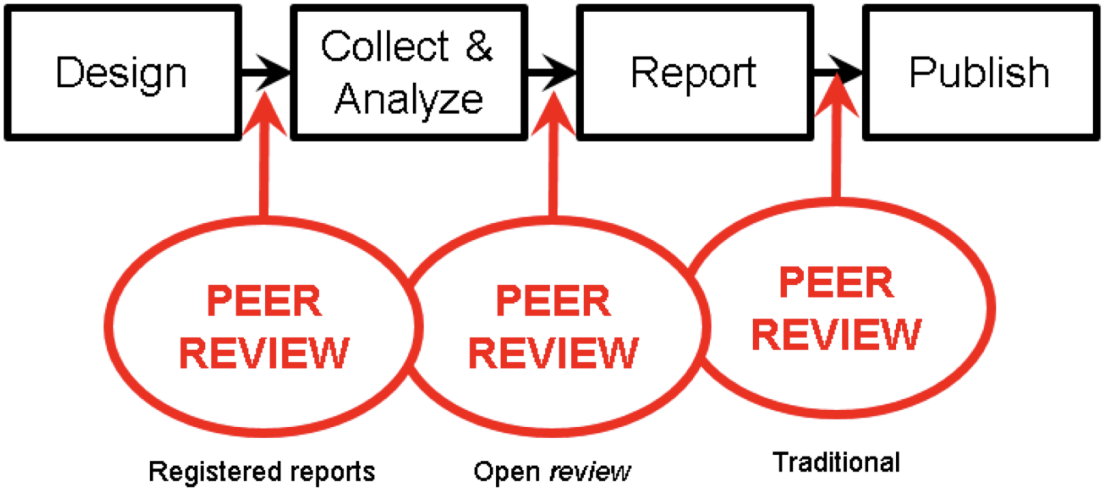
(f) Publishing **Null Results**: Researchers should be as compelled to publish negative or null findings as positive findings.

i. Link example 1

ii. Link example 2

[1] *embrace* being a data parasite.

(g) **Other (a)venues for publication**: Be aware that there are other avenues for publication, besides post-final-report peer review, such as registered reports and open peer review.



(h) Publication *Following* **Pre-Registration**: If completed and accepted during design phase, the full article can then be submitted for a second round of reviews. The article cannot be rejected at this point due to null findings but can be rejected for problems with the study itself (i.e. poor retention of study participants yielding insufficient power).

2 Precision Medicine and "-Omics"

2.1 General Concepts

Genetics Home Reference from the National Library of Medicine[2] (Link to online primer, the pdf file is available, 228 pages).

- Cells and DNA
- Mutations and Health
- How Genes Work
- Gene Families
- Inheriting Genetic Conditions
- Genetics and Human Traits
- Genetic Consultation
- Genetic Testing
- Newborn Screening
- Gene Therapy
- The Human Genome Project
- Genomic Research
- Precision Medicine

2.2 Key Concepts

Genomics, pharmacogenomics, (bioinformatics in genetic analysis), personalized care.

[2]National Library of Medicine. Genetic Home Reference - A service of the US National Library. 2017; https://ghr.nlm.nih.gov/primer. Accessed January 4, 2017.

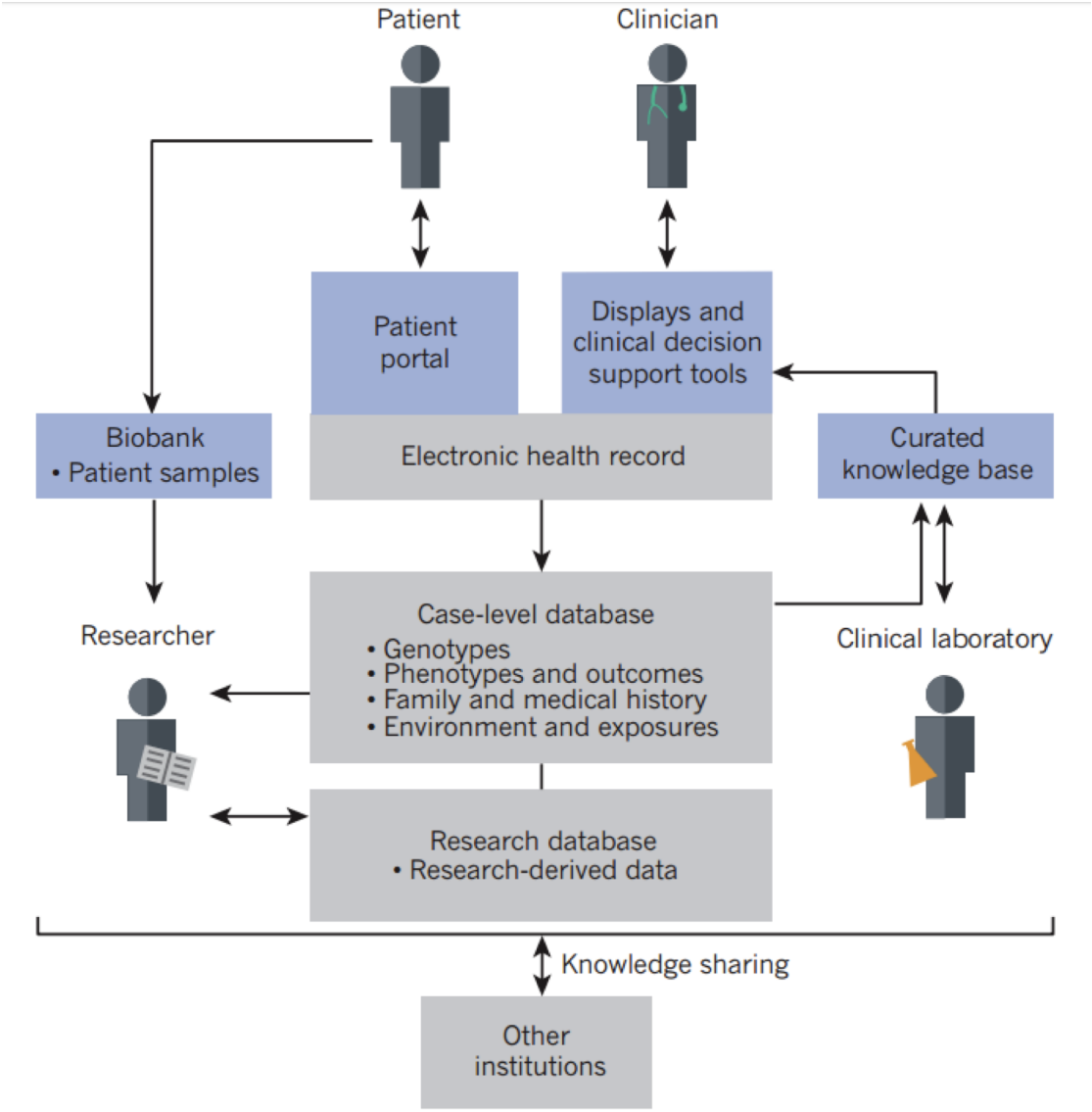
2.3 Definitions

- **Precision Medicine**: According to the National Institutes of Health (NIH), precision medicine is defined as "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person."
- **Genomics vs. Genetics**: According to the WHO, Genetics is the study of heredity while Genomics is the study of genes and the relationships between genes.
- **Pharmacogenomics**: According to the National Library of Medicine, "Pharmacogenomics is the study of how genes affect a person's response to drugs." (combines pharmacology-the science of drugs and genomics-the study of genes and their functions).
- **Bioinformatics**: The discipline that applies the information technology, computer science, math to facilitate the knowledge generation of genetics/genomics, molecular biology, and proteomics.[3]
- **Personalized care**: similar to the Precision Medicine; not necessary to focus on "genes, environment, and lifestyle."

[3]Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods of Information in Medicine. 2001;40(4):346-358.

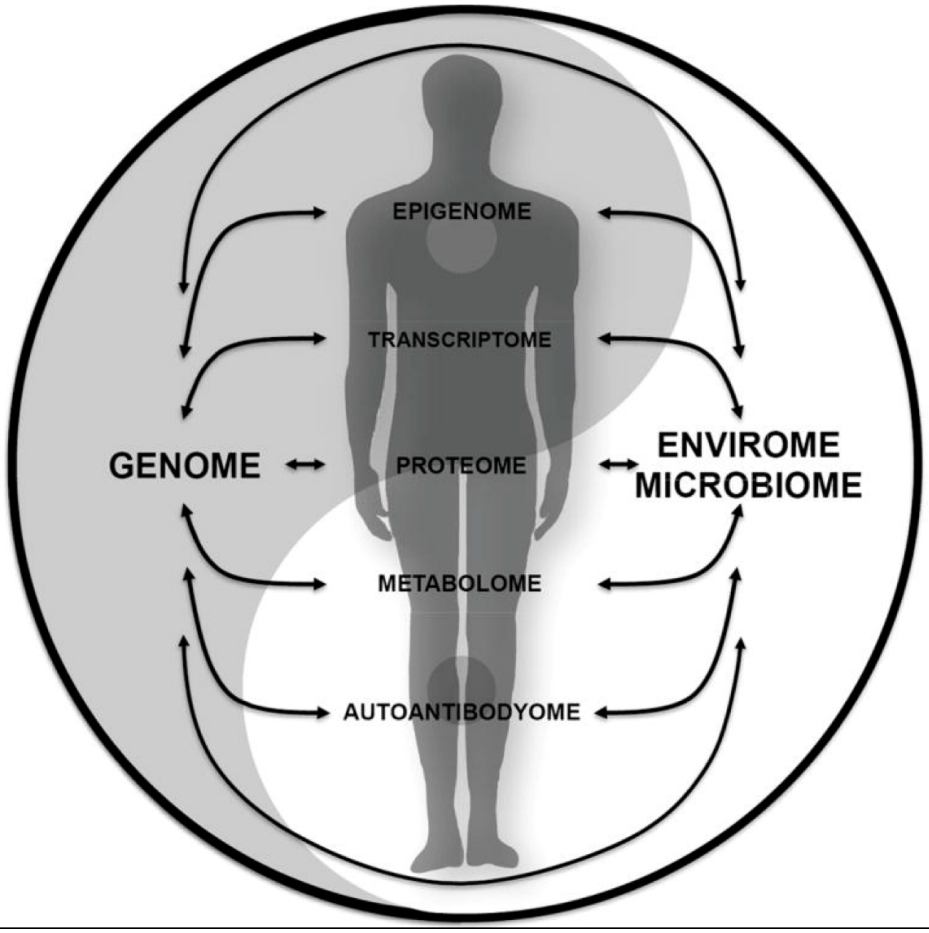
2.4 Practical Application of Precision Medicine and -omics within Healthcare System

"Genomics in Precision Medicine" from (Aronson & Rehm, 2015).[4]



[4]Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature. 2015;526(7573):336-342.

The concept of "Personal Omics Profile."[5]



[5]Chen R, Snyder M. Promise of personalized omics to precision medicine. Wiley Interdisciplinary Reviews. Systems Biology and Medicine. 2013;5(1):73-82.

2.5 Practical Applications of Precision Medicine and -omics by Point of Care

Summary of Nursing Implications of Personalized and Precision **Oncology** Care.[6]

TABLE 2.
Summary of Nursing Implications of Personalized and Precision Oncology Care

Patient Assessment and Management

Pre-diagnosis

- Assess clinical risk factors
- Screening as appropriate (family history, clinical risk factor testing, environmental risks)
- Explain genetic risk testing (implications, results, validity, familial implications)

Diagnostics

- Discuss genetic testing for relevant mutations associated with clinical presentation
- Discuss tumor testing for identification of somatic mutations

Treatment/Prognosis

- Identify patients for whom any 'omics or genetic testing platforms are appropriate for guidance in treatment considerations or prognostication
- Discuss recommended therapies based on tumor type and mutation identification
- Use clinical decision support tools to integrate personalized approaches and patient data (biomarker, patient-reported data, clinical data) into patient treatment plan discussions

Ongoing Monitoring and Management

- Support patient in determining and following through on family implications for genetic test results
- Psychosocial support for patient throughout chronic phase of oncology care

Implications for Nursing Education

- Understanding of and ability to explain genetic and genomic tests, their validity and the meaning of results that will be obtained
- Navigation of the ethical, legal and social issues involved in genetic and genomic testing
- Awareness of referral base for genetic counseling, clinical trials, or specialty care

[6] Vorderstrasse AA, Hammer MJ, Dungan JR. Nursing implications of personalized and precision medicine. Seminars in Oncology Nursing. 2014;30(2):130-136.

Nursing Genomic Science Blueprint Mapped to National Institute of Nursing Research (NINR) Strategic Plan Areas[7] (part of the table).

Table 1. Nursing Genomic Science Blueprint Mapped to National Institute of Nursing Research (NINR) Strategic Plan Areas

NINR strategic plan areas	Specific nursing research categories	Advisory panel genomic nursing research topic areas[a]
Health promotion and disease prevention	Risk assessment	a. Biologic plausibility (e.g., pathways, mechanisms, biomarkers, epigenetics, genotoxicity) b. Comprehensive screening opportunities (e.g., family history, identify risk level [population-based average and elevated]) c. Components of risk assessment (e.g., biomarkers, family history) d. Risk-specific healthcare decision making
	Communication	a. Risk communication (e.g., interpretation, timing, risk reports to the healthcare provider and client[b]) b. Informed consent c. Direct-to-consumer marketing and testing (e.g., uptake, utilization, dissemination)
	Decision support	a. Informed consent b. Match of values/preferences with decision made c. Risk perception/risk accuracy d. Effect of decision support on decision quality (e.g., knowledge, personal utility)
Advancing the quality of life	Family	a. Family context (e.g., family functioning, and structure, family relationships, and communication) b. Ethical issues c. Healthcare provider communication with families
	Symptom management	a. Biologic plausibility (e.g., pathways, mechanisms, biomarkers, epigenetics) b. Clinical utility c. Personal utility d. Pharmacogenomics (e.g., therapy selection, medication titration) e. Decision making f. Evidence-based effectiveness of approaches
	Disease states (encompassing acute, common complex and chronic)	a. Genomic-based interventions that reduce morbidity and mortality b. Gene/environment interactions (e.g., epigenetics, genotoxicity) c. Pharmacogenomics d. Evidence-based effectiveness of treatments/support
	Client self-management	a. Collecting and conveying information that informs self management (e.g., family history) b. Lifestyle behaviors c. Environmental exposure and protection (e.g., occupational) d. Synergy of client and provider expectations (e.g., client/family centered care) e. Personal utility

[7]Genomic Nursing State of the Science Advisory P, Calzone KA, Jenkins J, et al. A blueprint for genomic nursing science. Journal of Nursing Scholarship. 2013;45(1):96-104.