Biological Research for Nursing
2017, Vol. 19(1) 18-27
© The Author(s) 2016
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1099800416666716
journals.sagepub.com/home/brn

Symptom Science: Repurposing Existing Omics Data

Nicole D. Osier, BSN, BS, RN, PhD Student[1], Christopher C. Imes, PhD, RN[1], Heba Khalil, RN, PhD[2], Jamie Zelazny, MPH, RN, PhD Student[1], Ann E. Johansson, BSN, RN, PhD Student[1], and Yvette P. Conley, PhD[1]

Abstract

Omics approaches, including genomics, transcriptomics, proteomics, epigenomics, microbiomics, and metabolomics, generate large data sets. Once they have been used to address initial study aims, these large data sets are extremely valuable to the greater research community for ancillary investigations. Repurposing available omics data sets provides data to address research questions, generate and test hypotheses, replicate findings, and conduct mega-analyses. Many well-characterized, longitudinal, epidemiological studies collected extensive phenotype data related to symptom occurrence and severity. While the main phenotype of interest for many of these studies was often not symptom related, these data were collected to better understand the primary phenotype of interest. A search for symptom data (i.e., cognitive impairment, fatigue, gastrointestinal distress/nausea, sleep, and pain) in the database of genotypes and phenotypes (dbGaP) revealed many studies that collected symptom and omics data. There is thus a real possibility for nurse scientists to be able to look at symptom data over time from thousands of individuals and use omics data to identify key biological underpinnings that account for the development and severity of symptoms without recruiting participants or generating any new data. The purpose of this article is to introduce the reader to resources that provide omics data to the research community for repurposing, provide guidance on using these databases, and encourage the use of these data to move symptom science forward.

Keywords

omics, big data, symptoms, repurposing

Omics research approaches generate large data sets, a type of "big data" that is used to test biological pathways of interest as well as perform nonparametric evaluations that identify what regions of the genome, genes, or biological pathways are related to the phenotype of interest. The practice of making omics data available to the research community for repurposing is gaining momentum, and many funding agencies mandate the sharing of these data. An example is the National Institutes of Health (NIH) Genomic Data Sharing Policy, which promotes responsible sharing of any large-scale genomic data (genomic, transcriptomic, epigenomic, etc.) generated with NIH funding (https://gds.nih.gov). Omics data from large, longitudinal, well-characterized cohorts, such as the Framingham Heart Study, the Nurses' Health Study, and the Jackson Heart Study, are available for repurposing by the research community for further scientific inquiry. Repurposing available omics data sets can provide data to address research questions, generate or test specific hypotheses, and conduct mega-analyses, all of which can be used to aid in directing a program of research.

The National Institute of Nursing Research (NINR) is committed to genomic nursing research (Tully & Grady, 2015), and based on *The Blueprint for Genomic Nursing Science* (Calzone et al., 2013) and the NINR Four Key Research Themes (NINR, 2016), there are many instances where repurposing existing omics data could be used to address NINR's mission. For example, one research theme that would benefit from repurposing existing omics data is symptom science and the promotion of precision health strategies. Many of the well-characterized, longitudinal, epidemiological studies that have made their omics data available collected extensive phenotype data related to symptom occurrence and severity, endogenous and exogenous exposures, and comorbidities in addition to extensive demographic data. The main phenotype of interest for many of these studies was often not symptom related, but

[1] School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA
[2] School of Nursing, Applied Science University, Amman, Jordan

Corresponding Author:
Yvette P. Conley, PhD, School of Nursing, University of Pittsburgh, 3500 Victoria Street, 440 Victoria Building, Pittsburgh, PA 15261, USA.
Email: yconley@pitt.edu

data on symptoms were collected to better understand the primary phenotype of interest. Indeed, the symptom-related data were often included to address potential confounding variables. These data could be instrumental in developing precision strategies for the prevention and treatment of symptoms.

An additional area that is receiving a considerable amount of effort is the linking of electronic health records to omics data for research purposes. Examples of these initiatives include the Electronic Medical Records and Genomics (*n* ~ 18,600) and the Resource of Genetic Epidemiology Research on Aging (*n* ~ 78,400) projects. Nurse scientists are ideal investigators for these initiatives and should be at the front line of efforts that leverage big data–related opportunities to improve patient care (Brennan & Bakken, 2015).

The purpose of this article is 3-fold: (1) to introduce the reader to resources that provide omics data to the research community for repurposing, (2) to provide guidance on the utility of these resources, and (3) to encourage the use of these available data sources to move symptom science forward.

Existing Omics Data Available to the Research Community

The suffix *omics*, when used in terms such as *genomics, transcriptomics, proteomics, epigenomics, microbiomics, and metabolomics*, indicates looking at the totality of data for the chosen approach. The term genomics, therefore, refers to characterizing the DNA variability in not one gene, but all genes across the genome. The term transcriptomics refers to characterizing not one RNA transcript, but all RNA transcripts from a cell, tissue, or organism.

The Human Genome Project (HGP) was the first big data omics project in which the resulting data were made readily available to the larger scientific community (National Human Genome Research Institute, 2016). The sequence of the 3 billion DNA bases in the human genome revealed more than a decade ago through the HGP is available through a variety of integrated databases. The pace of the collecting of omic-type data is showing no sign of slowing down as the research community increasingly recognize the promise of omics. Moreover, the techniques for collecting these data are rapidly improving. For example, sequencing the human genome for the HGP took over a decade, while next generation sequencing can sequence a human genome in days and at a fraction of the cost (Conley et al., 2013).

Table 1 provides information about selected databases and resources available to the research community for repurposing of omics data. These databases are continuously being updated and are curated by reliable agencies; therefore, they represent a means to stay up-to-date on available resources. Additionally, we direct readers to the 2016 Database Issue of *Nucleic Acids Research* (Rigden, Fernandez-Suarez, & Galperin, 2016), which provides an overview of a variety of omics databases. This issue is freely available online at http://nar.oxfordjournals.org. A more thorough list of databases available online can be found at http://oxfordjournals.org/nar/database/c/

Guidance on Using Omics Databases

Conceptualizing how to use the data available through omics databases to address symptom science starts by identifying the problem. For example, we do not fully understand the biological underpinnings of fatigue or variability in those who experience fatigue. This knowledge could provide the evidence base to guide investigations into the development of new interventions and aid in progress toward precision health care for fatigue. Two possible solutions to address this problem are as follows: (1) recruit, phenotype, and collect omics data from biospecimens for thousands of subjects and conduct a nonparametric analysis of these data to determine what genes and biological pathways are involved in fatigue or (2) obtain already existing omics and phenotype data from thousands of subjects through omics databases and conduct a nonparametric analysis of these data to determine what genes and biological pathways are involved in fatigue. Both of these potential solutions would result in biologically guided hypothesis generation and establish the preliminary data to support further investigation; however, the second solution efficiently addresses the study question using existing resources. In addition to its utility for hypothesis generation, as in the example above, data from omics databases can be used for hypothesis testing, including analysis of targeted candidate biological pathways.

These omics databases are catalogs of studies that include key information about the studies, themselves, and the omics data available. Users can conduct searches using key words such as the phenotype of interest (e.g., *depression, anxiety, pain,* or *fatigue*) or may search the databases for specific studies or cohorts. Using the database of genotypes and phenotypes (dbGaP) as an example, a search for the key word *fatigue* would return not only studies that collected fatigue-specific data but also those that mention fatigue in other ways that may not be as relevant, such as if a subject stopped data collection due to fatigue. Users can limit their searches in dbGaP by clicking on the "limit" hyperlink and choosing from a long list of fields to search, including, for example, "variable." dbGaP also has an advanced search function that allows users to build a search using Boolean terms like *and*, *or*, and *not* to build specific searches across multiple fields. The dbGaP handbook provides further instruction on optimizing searches within dbGaP (www.ncbi.nlm.nih.gov/books/NBK154410).

Access to some of these databases is restricted and requires an application outlining the intended use, a data sharing agreement, and in some cases approval by a local institutional review board. For example, as part of the application process to gain access to data in dbGaP, database administrators conduct a review that evaluates the proposed research and terms of the original consent given by the research subjects. The procedure to secure access to dbGaP data can be found at https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login. In addition, dbGaP recently opened access to a collection of studies with data appropriate for general research use that researchers can use for

Table 1. Selected Resources Available Online for the Repurposing of Omics Data.

Resource	URL	Description
Database of genotypes and phenotypes (dbGaP)	http://www.ncbi.nlm.nih.gov/gap/	Searchable database of studies that have generated genotype data, usually through GWAS or sequencing-based studies. Also searchable for phenotype variables connected to the studies
GWAS central	www.gwascentral.org	(Formerly called the Human Genome Variation Database of Genotype-to-Phenotype Information.) Database of summary-level findings from genetic association studies, both large and small
Multi Omics Profiling Expression Database (MOPED)	https://www.proteinspire.org/MOPED	A growing multiomics resource that supports rapid browsing of expression information from publicly available studies on model organisms and humans. MOPED is designed to simplify the comparison and sharing of proteomics data for the greater research community
Gene Expression Omnibus (GEO)	www.ncbi.nlm.nih.gov/geo	Public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomic data submitted by the scientific community. In addition to data storage, a collection of web-based interfaces and applications are available for users to query and download the studies and gene-expression patterns that are stored in GEO
European Bioinformatics Institute (EMBL-EBI)	www.ebi.ac.uk	Freely available and up-to-date molecular databases including those that let investigators share data from life-science experiments
Wellcome Trust Case Control Consortium (WTCCC)	www.wtccc.org.uk	The primary purpose of the WTCCC is to accelerate efforts to identify genome sequence variants influencing major causes of human morbidity and mortality through implementation and analysis of large-scale GWAS. Additional objectives include the development and validation of informatics and analytical solutions appropriate to the scale and nature of the project as well as use of the data generated to answer important methodological and biological questions relevant to association studies in general and in the UK in particular (e.g., issues of population substructure). The consortium anticipates that data generated from the project will be used by others, with potential applications including developing new analytical methods, understanding patterns of polymorphisms, and guiding selection of markers to map genes involved in specific diseases. Access to summary data and individual-level genotype data is available by application to the WTCCC Data Access Committee
The NIMH Repository and Genomics Resource	https://www.nimhgenetics.org/available_data/data_biosamples/gwas_data.php	One of its missions is to receive and process clinical and genetic data and to distribute it to approved investigators
Gemma	http://chibi.ubc.ca/Gemma	Database, analysis software system, and website for genomics data reuse and meta-analysis
GeneProf	http://www.geneprof.org	Open web resource for analyzed functional genomics experiments, mostly reanalysis of publicly available RNA sequencing (RNA-seq) and chromatin immunoprecipitation followed by sequencing (ChIP-seq) data sets
CistromeFinder	http://cistrome.org/finder	Data portal to assist with the query, evaluation, and visualization of publicly available ChIP-seq and DNase I hypersensitive sites sequencing (DNase-seq) data. It is integrated with the UCSC genome browser
1,000 Genomes	http://www.1000genomes.org/category/data-reuse	Publicly available data from sequencing the genome of individuals. Data available include DNA variant calls, aligned sequence data, and raw sequence data
Phenotype-Genotype Integrator (PheGenI)	http://www.ncbi.nlm.nih.gov/gap/phegeni	Phenotype-oriented resource, intended for clinicians and epidemiologists interested in following up results from GWAS, can facilitate prioritization of variants to follow-up, study design considerations, and generation of biological hypotheses

Note. GWAS = genome-wide association studies.

secondary data analyses without limitation. See the NCBI dbGaP collection's website (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/collection.cgi?study_id=phs000688.v1.p1) for more information about these data. The larger dbGaP database houses information on all studies available for data repurposing (www.ncbi.nlm.nih.gov/gap).

The initial search results in dbGaP are presented in a tabular format that lists study name, whether the data are currently available or embargoed (along with release date), number of participants with data available through dbGaP, type of study (case–control, family based, cohort, longitudinal, etc.), and the omics data collection platform(s) used. The study name is clickable and links you to more detailed information about the study including descriptions of the study, study design, phenotyping, and omics data collection efforts; inclusion and exclusion criteria for recruitment; how and when the biological samples were collected; and information about the study history. This information includes selected publications related to the study, identification (requestor name, affiliation, project title, and approval date) of researchers already granted access to the data, and names and affiliations of the investigators involved in the original study. These initial search results are a great starting place for learning more about what data are available for your phenotype of interest.

Exemplar Database Searches

To see how well represented symptom-related data might be in publicly available omics databases, we conducted some exemplary searches using mostly dbGaP and focused primarily on DNA variability, gene expression, and DNA methylation data. These searches reflect a cross-sectional snapshot of extremely dynamic databases to which data and information are frequently added. Therefore, readers interested in these symptoms are encouraged to conduct up-to-date searches of these databases. We conducted searches related to the following phenotypes: cognitive impairment, fatigue, gastrointestinal (GI) distress/nausea, sleep, and pain. We searched specifically for studies that included one or more of these symptoms as a study variable measured in conjunction with a variety of omics data. Table 2 details the steps we took to conduct these searches.

Through our search, we found that cognitive impairment was measured as a variable in 19 studies with omics data available, fatigue in 13 studies, GI distress/nausea in 16 studies, sleep in 30 studies, and pain in 29 studies. Some of the studies focused on specific patient populations, such as those with diabetes, cardiovascular disease, chronic obstructive pulmonary disease, stroke, and a variety of mental health and neurological disorders. Fatigue was the only symptom for which study subjects seemed to be limited to the adult population; studies in children were available for the other symptoms. One issue that became apparent through the course of our search was that phenotypic data were not collected in a consistent manner between studies. Of particular interest are the studies that collected omics data and data on these symptoms over time in healthy, community-dwelling individuals, such as the Jackson Heart Study (n ∼ 3,300), the Women's Health Initiative (n ∼ 64,000), the Cardiovascular Health Study (n ∼ 5,200), and the Framingham Cohort studies (n ∼ 15,000). Table 3 presents more information about the search results, including instruments and measurements the studies used to collect symptom phenotypes and demographic data for the study populations.

Common Challenges With and Recommendations for Using Shared Omics Data

There are many challenges associated with using shared omics data that can impact the utility and interpretation of the data. Common examples include the following: data in the database represent only a subset of subjects from the study; lack of in-depth information on how symptom phenotypes were collected and adjudicated; and important methodological information about the study, for example, recruitment schema, is not available through the database.

It is not uncommon for data from a subset of available subjects from a cohort to be deposited into a data sharing database. Investigators may only make data available through omics database requests from subjects who agreed to release of their data to other investigators not affiliated with the original project or the original investigators. This means that, for some studies, a substantial number of subjects' omics data cannot be made available through the database. However, subjects may have consented to sharing data in ways other than through publicly available databases. For example, they may have consented to sharing of data with direct collaborators of the original investigators. Therefore, collaborating directly with the study investigators for data access may result in access to more subjects, leading to increased power and potentially reducing issues around selection bias.

A lack of understanding of how symptom phenotype data were generated in a study can impact authors' abilities to effectively combine phenotype data for mega-analyses and to assess the rigor and robustness of the data. Many of the studies that have data deposited into the databases have websites established for those studies, for example, the Cardiovascular Health Study (CHS) at https://chs-nhlbi.org. The study websites frequently provide better insight into the phenotypes collected than the data sharing databases and often have the data collection instruments available for review and provide the standardized protocols for data collection. Additionally, the study website may provide information about phenotypes not deposited into the database. For example, investigators collected data on quality of life, social support, and stressful life events in the CHS, but these phenotypes are not mentioned in dbGaP. The study website will also often have information about approved ancillary studies, including those being conducted by the study investigators that would not be represented in databases like dbGaP; thus, researchers can confirm that what they want to do has not been proposed already.

Making decisions about what data to request access to and then interpreting findings from the data analyses require in-depth information about the methodologies investigators used

Table 2. Details of Exemplar Database Searches for Data Related to Specific Symptoms.

Symptom and Steps	Rationale	Results
Cognitive impairment		
1. Searched dbGaP using the term *cognitive impairment*	Cognitive impairment was the symptom of interest	Search results: 153 variables, 0 analyses, 138 documents, and 20 data sets in 34 studies
2. Clicked on each study sequentially	To understand the overall purpose and aim of each study, the study population and why the symptom was included in the data set	In several studies, cognitive impairment was listed as an exclusion criterion for the study but wasn't actually measured as a variable of interest
3. Clicked on "Variables" tab	To review the variables collected pertaining to cognitive impairment	Variable folders appear on the right side of the page. Clicked through each folder to find the variable(s) relevant to cognitive impairment
4. Clicked on "Documents" tab	To review the assessments used to measure cognitive impairment	Reviewed data collection forms. Assessment of cognitive function varied from a Yes/No item completed on and Inclusion/Exclusion checklist to questionnaires assessing cognitive function to extensive neuropsychological assessment batteries
5. Tabulated findings into a table	To summarize and synthesize search results	Of the 34 studies identified in the search, 19 studies were appropriate
Fatigue		
1. Searched dbGaP using the term *fatigue*	Fatigue was the symptom of interest	Search results: 80 variables, 0 analyses, 130 documents, and 0 data sets in 41 studies
2. Clicked on the first study in the results, "phs000796.v1.p1 Genome Wide Association Study of Chronic TMD: Discovery Phase"	To learn the details of the study	Viewed the description of the study, the variables collected, study document provided, data sets, and molecular data
3. Clicked on the "Variables" tab	To search through the variables to determine if and how fatigue was measured	5 folders containing the variables collected as part of the study became available for searching
4. Clicked on "Phenotype—Details" folder	To discover how fatigue was measured	2 subfolders containing the variables collected as part of the study available for searching
5. Clicked on "Medical History" subfolder	To discover how fatigue was measured	14 subfolders containing data from study questionnaires available for searching
6. Clicked on "Short-Form (SF12)" subfolder	To discover how fatigue was measured	2 subfolders containing the variables from the SF12 available for searching
7. Clicked on "Physiological Measurements and Observations" subfolder	To discover how fatigue was measured	10 items available, including "Lot of Energy"
8. Clicked on "Lots of Energy"	To determine whether "Lots of Energy" could be considered a measure of fatigue	The item has scoring (*some of the time, a little of the time,* and *none of the time*) that could be used as a measure of fatigue
9. Went back to the "Phenotype—Details" subfolder and clicked on the "Psychological and Psychiatric Observations" subfolder	To search for additional measures of fatigue	11 subfolders containing data from study questionnaires available for searching
10. Clicked on "Profile of Mood States-Bipolar (POMS)" subfolder	To search for additional measures of fatigue	65 items available, including "Fatigued"
11. Clicked on "Fatigued"	To determine whether "Fatigued" could be considered a measure of fatigue	The item has scoring that could be used as a measure of fatigue
12. Went back to the "Phenotype—Details" subfolder and clicked on the "Psychological and Psychiatric Observations" subfolder	To search for additional measures of fatigue	11 subfolders containing data from study questionnaires available for searching
13. Clicked on "Symptom Check List 90 (SCL90)" subfolder	To search for additional measures of fatigue	90 items available, including "Low Energy"
14. Clicked on "Low Energy"	To determine whether "Low Energy" could be considered a measure of fatigue	The item has scoring that could be used as a measure of fatigue
15. Returned to original search, clicked on remainder of studies, and conducted the same steps for each	To learn the details of each study	Of the 41 studies identified in the search, 13 studies were appropriate

(continued)

Table 2. (continued)

Symptom and Steps	Rationale	Results
GI distress/nausea		
1. Searched dbGaP using the term *nausea*	To evaluate which studies and what types of studies included the key word *nausea*	Results included 44 studies that were a mix of case–control, cohort, family studies, and clinical trials; 79 variables measuring nausea and descriptions of how the variables were measured (e.g., anorexia, nausea, or vomiting in the past week)
2. Conducted advanced search with *nausea* in variable name	To narrow down studies to those with nausea as a variable	Results included 16 studies: 7 case–control and 8 longitudinal cohort or clinical trial
3. Repeated simple search with key words *GI distress* OR *gastric distress*	To evaluate which studies and what types of studies included the key words *GI distress* or *gastric distress*	Results included one study that mentioned *GI distress* in the background of the protocol but did not have a variable measuring it; no results for *gastric distress*
Sleep		
1. Used dbGaP to perform a search without limits (all fields) for the key word *sleep*	Using the term *sleep* in its original form was a logical starting point for the initial search	Results included 62 studies; however, not all the retrieved study seemed germane to the topic of sleep
2. Used dbGaP to perform a search with limits (variable field only) for the key word *sleep*	To identify additional studies that included sleep-related variables in the data collected from participants	Results included 21 studies that were added to the table; however, some of the retrieved studies included a single sleep question on an assessment of another symptom (e.g., depression, anxiety)
3. Used dbGaP to perform a search with limits (variable field only) for the key words *sleep disorders* OR *apnea* OR *restless leg* OR *sleep walk* OR *dream* OR *REM* OR *circadian* OR *bedwetting* OR *night terrors*	To identify relevant studies that collected data on other sleep-related phenotypes or one or more aspect of the sleep experience by broadening the search terms	Results included numerous (over 50) studies; however, many had been previously identified in the unlimited *sleep* search described in Step 1. In total, 9 new studies were retrieved by broadening search terms and subsequently added to the table
4. Used PhenGenI to perform a basic search by phenotype selection for the key word *sleep*	To identify genetic factors related to the phenotype of sleep	282 associations, 18 genes, and 12 SNPs were identified; however, no data were found in available databases (eQTL and dbGAP). Thus, no new studies were added to the table
Pain		
1. Searched dbGap database using *pain* as key word	Pain was the symptom of interest	A total of 105 studies were identified; however, a number of the studies were not directly related to pain
2. Conducted an advanced search in dbGaP using *pain* in the "variable" field	To limit search to studies that collected pain as a variable	A total of 33 studies were identified
3. Conducted an advanced search in dbGaP using *pain* in the "variable description" field	To limit search to studies that collected pain as a variable	A total of 31 studies were identified
4. Conducted an advanced search using *pain* in the "variable name" field	To limit search to studies that collected pain as a variable	A total of 8 studies were identified
5. Conducted an advanced search using *pain* in the "study name" field	To limit search to studies that collected pain as a main focus	A total of 2 studies were identified
6. Clicked on each study and then clicked through each pain variable to look at details	To make sure that pain was included as a major outcome in the selected studies	Only 29 studies had data available directly related to pain

Note. eQTL = expression quantitative trait loci browser; dbGaP = database of Genotypes and Phenotypes; PhenGenI = Phenotype-Genotype Integrator; SNPs = single nucleotide polymorphisms; GI = gastrointestinal.

in the original study, including recruitment strategies and inclusion and exclusion criteria. How subjects were recruited and selected are important considerations when assessing the quality of the original study, determining how appropriate the study data would be for addressing the research questions or hypotheses of another study, and interpreting the findings from a study that utilized the data from the original study. Fortunately, many of the researchers who have made their data available in these databases have published articles about their methodologies; therefore, a thorough literature search can capture this information. Other information that can be extracted from the literature may include the number of subjects in the original study, which can differ significantly from what is available through the databases, how certain types of data were collected, and what has been published so far (to prevent overlap of effort and highlight opportunities for additional inquiry).

Table 3. Exemplar Results for Symptom-Based Searches in Omics Databases.

Symptom	# of Studies	Clinical Characteristics of Study Populations	Demographic Characteristics of Study Populations	Measurements/Instruments Used to Measure the Symptom	Omics Approach Used
Cognitive impairment	19	Healthy community-dwelling participants; patients with cardiovascular disease, diabetes, Huntington's disease, Alzheimer's disease, ophthalmic conditions, dementia, ADHD, multiple sclerosis, bipolar disorder, ischemic stroke, and Parkinson's disease; samples designed to focus on women's health and aging	Predominantly Caucasian and African American; males and females represented, though some studies did focus solely on women (e.g., the Women's Health Initiative); Ages 6 years and up; primarily family-based and unrelated cohort studies	Mini-Mental State Exam; telephone interview for cognitive status; neuropsychological testing batteries including subtests from the Wechsler Adult Intelligence and Memory Scales, Boston Embedded Figures Test, Trailmaking Test parts A & B, Grooved Pegboard test, Halstead Finger-tapping Test, Star Drawing Test; cognitive assessment interviews including Delayed Word Recall, Digit Symbol Substitution, Word Fluency, Stroop Color Word Test, Facial Recognition Test, Smell Identification Test, Dual Verbal Working Memory Test, Hopkins Verbal Learning Test Revised, Immediate and Delayed Recall Tests, Serial Response Task, Tower Task, Emotion Recognition Test, WAIS-III, Simple/Choice Reaction Time tasks, American Adult National Reading Test, Set-Shifting Task, Verbal Fluency Buttons Task; Penn Conditional Exclusion Test, Penn Continuous, Performance Test, Letter N-Back Test, Penn Word Memory Test, Penn Face Memory Test, Visual Object Learning Test, Penn Verbal Reasoning, Penn Line Orientation Test, Facial Emotion Processing; Sensory Motor Processing Speed, Logical Memory IA, Digit Span Forward, Digit Span Backward, Category Fluency: Animals and Vegetables, Digit Ordering; Logical Memory IIA, Logical Memory IIA-Delayed	GWAS, exome sequencing, whole genome gene expression, miRNA, whole genome DNA methylation
Fatigue	13	Healthy community-dwelling participants; patients with cardiovascular disease, COPD, Type 1 diabetes, major depressive disorder, benign ethnic neutropenia/leukopenia, multiple sclerosis, amphetamine addiction, and TMJD; samples designed to focus on women's health	Mostly Caucasians except for the NHLBI Cleveland Family Study Candidate Gene Association Resource (44% AA) and the Benign Ethnic Neutropenia/Leukopenia in African Americans (all AA); good balance of males and females; ages range from young adults (18–35 years) to older adults (mean age of 72 years)	SF-36: Vitality (energy/fatigue) subscale (*n* = 6); SF-12 (1 vitality item; *n* = 2); Profile of Mood States (*n* = 2); Maastricht Vital Exhaustion Questionnaire, 21-item version (*n* = 1); Quick Inventory of Depressive Symptomatology (1 fatigue item; *n* = 1); Functional Assessment of Multiple Sclerosis (fatigue items; n = 1)	GWAS, exome sequencing

(continued)

Table 3. (continued)

Symptom	# of Studies	Clinical Characteristics of Study Populations	Demographic Characteristics of Study Populations	Measurements/Instruments Used to Measure the Symptom	Omics Approach Used
GI distress/ nausea	16	Healthy community-dwelling participants; patients with cardiovascular disease, Parkinson's disease, Alzheimer's disease, cancer, TMJD, multiple sclerosis, cataracts, bipolar disease, addiction, diabetes, stroke, Crohn's disease, neurodevelopmental conditions, COPD, asthma, obesity, developmental/ intellectual delay, congenital anomaly; samples designed to focus on women's health, pharmacogenomics, oral health, children's health	Males and females; ages 2 to >65 years; Caucasian, African American, Asian; case–control and cohort	Questionnaires (Pennebaker Inventory of Limbic Languidness, Symptom Checklist 90, Computerized Neurocognitive Test Battery), medical examiner's report, clinical exam	GWAS, whole genome sequencing, whole exome sequencing, whole genome genotyping, SNP arrays, 16 S rRNA, metagenomic sequencing, FISH, oligo CGH
Sleep	30	Healthy community-dwelling participants; patients with cardiovascular disease, chronic TMJD, diabetes, COPD, preterm birth, ophthalmic conditions, complex pediatric disorders, Alzheimer's disease, major depression, bipolar disorder, alcoholism	Pediatric through elderly; some focused only on African Americans, others were more diverse but ended up being mostly Caucasian; some were limited to women only (e.g., Women's Health Initiative) while most included males and females	Self-reported sleep data assessed via interview or researcher-generated form: sleep duration (hr), sleeping pill use (yes/no), if sleeping pills were used in the past 4 weeks (never, <1/week, 1–2/week, 3–4/week, 5+/week), feeling sleepy during day in the past 4 weeks (never, <1/week, 1–2/week, 3–4/week, 5+/week), sleep disturbances in the past week (yes/no), sleep restlessness (none/rarely, some or a little of the time, a moderate amount of time, most of the time), sleep restlessness in the past week (<1 day, 1–2 days, 3–4 days, 5–7 days), sleep on 2+ pillows to help breathe (yes/no/do not know), sleepy in the day time (yes/no), trouble falling asleep (yes/no), oxygen use during sleep (yes/no), awakened by cough (yes/no), awakened by shortness of breath/tightness in chest (yes/no), sleep apnea ever (yes/no/do not know), sleep apnea diagnosed by doctor (yes/no), sleep apnea age of onset, sleep apnea treatment in the past 12 months (yes/no) Self-report of sleep data assessed as part of established measures/tools: Inventory of Depression Severity (Items on falling asleep, sleep during the night, waking up too early, sleeping too much), Amsterdamse Biografische Vragenlijst (neuroticism item: Are you often unable to sleep because there are so many thoughts going through your head?)	GWAS, exome sequencing, whole genome sequencing, DNA methylation (targeted and whole genome), global gene expression, miRNA expression

(continued)

Table 3. (continued)

Symptom	# of Studies	Clinical Characteristics of Study Populations	Demographic Characteristics of Study Populations	Measurements/Instruments Used to Measure the Symptom	Omics Approach Used
Pain	29	Healthy community-dwelling participants; patients with Sjogren's syndrome, Parkinson's disease, TMJD, cardiovascular disease, diabetes, COPD, stroke, ophthalmic conditions, melanoma, recurrent abdominal pain, Alzheimer's disease, multiple sclerosis, schizophrenia, major depression, bipolar disorder, lupus	Pediatric through elderly; some focused on specific populations (e.g., African Americans, Dutch, UK residents), while others were more diverse but ended up being mostly Caucasian; males and females represented	Mechanical cutaneous pain, aftersensation ratings for 256 mN probe (numeric rating), Comprehensive Pain and Symptom Questionnaire, Graded Chronic Pain Scale score (for facial and nonfacial pain), Pain Catastrophizing Scale, Coping Strategies Questionnaire–Revised, bodily pain index, ID Pain self-administered screening tool, inventory of depression severity, many questionnaires related to pain site and severity developed specifically for studies	GWAS, exome sequencing, whole genome sequencing, microbiome data

Note. AA = African American; ADHD = attention deficit hyperactivity disorder; CGH = comparative genomic hybridization; COPD = chronic obstructive pulmonary disease; DNA = deoxyribonucleic acid; FISH = fluorescence in situ hybridization; GI = gastrointestinal; GWAS = genome-wide association study; miRNA = micro ribonucleic acid; NHLBI = National Heart, Lung, and Blood Institute; rRNA = ribosomal ribonucleic acid; SF = Short Form survey; SNP = single nucleotide polymorphism; TMJD = temporo-mandibular joint disorder; WAIS III = Wechlser Adult Intelligence Scale III.

Other challenges associated with using shared omics data include adjudication of omics data, combining omics data collected using multiple data collection platforms, and the ethical and policy-related issues around sharing the data and subsequent findings. These are important issues currently receiving attention in the omics scientific community. The white paper from the National Consortium for Data Science entitled *Data to Discovery: Genomes to Health* (Ahalt et al., 2014) provides more information related to these challenges.

An additional challenge related to using data available through omics databases is that analyses of data and interpretation of findings, particularly if whole genome data will be analyzed, require specialized statistical and bioinformatics expertise. Although the entire research community faces this issue, it may be especially challenging for nurse scientists who may not have access to the necessary resources or statistical expertise among their network of current collaborators. We encourage scientists embarking on a project involving the access of data through omics databases to seek guidance from potential colleagues at bioinformatics and computational genomics cores, which are found at many universities and medical centers and whose services are often available to collaborators from outside of their university. These cores are charged with facilitating translational research; therefore, they are ideal collaborators for nurse scientists who conceptualize the translational use of these data.

Conclusion

Scientists today have a golden opportunity to address big questions and substantially move symptom science forward by taking advantage of the abundance of phenotype and omics-related data available for the asking. The wealth of data available and the infinite research questions that can be asked with these data are inspiring. Tremendous resources have been invested into the studies that are making their omics data available to the research community. Many of these studies recruited thousands of individuals and families, collected and banked biospecimens, and collected copious amounts of personal and phenotype data (often longitudinally) and a variety of omics data. Nurse scientists are well positioned to repurpose these data to answer new questions and generate and test new hypotheses related to symptoms that the research community has not been yet addressed.

Author Contribution

Nicole Osier contributed to conception and design, acquisition, analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be accountable for all aspects of work ensuring integrity and accuracy. Christopher Imes contributed to conception and design, acquisition, analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be accountable for all aspects of work ensuring integrity and accuracy. Heba Khalil contributed to conception and design, acquisition, analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be accountable for all aspects of work ensuring integrity and accuracy. Jamie Zelazny contributed to conception and design, acquisition, analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be accountable for all aspects of work ensuring integrity and accuracy. Ann Johansson contributed to conception and design, acquisition, analysis, and interpretation; critically revised the manuscript; gave final approval; and agrees to be accountable for all aspects of work ensuring integrity and accuracy. Yvette Conley contributed to conception and design, acquisition, analysis, and interpretation; drafted the manuscript; critically revised the manuscript; gave final approval; and agrees to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors acknowledge support from the National Institute of Nursing Research (T32NR009759 and F31NR014957).

References

Ahalt, S., Bizon, C., Evans, J., Erlich, Y., Ginsberg, G., Krishnamurthy, A., … Wilhelmsen, K. (2014). *Data to discovery: Genomes to health. A white paper from the National Consortium for Data Science*. Chapel Hill, NC: RENCI, University of North Carolina at Chapel Hill. doi:10.7921/G03X84K4

Brennan, P. F., & Bakken, S. (2015). Nursing needs big data and big data needs nursing. *Journal of Nursing Scholarship*, *47*, 477–484.

Calzone, K. A., Jenkins, J., Bakos, A. D., Cashion, A. K., Donaldson, N., Feero, W. G., … Webb, J. A. (2013). A blueprint for genomic nursing science. *Journal of Nursing Scholarship*, *45*, 96–104.

Conley, Y. P., Biesecker, L. G., Gonsalves, S., Merkle, C. J., Kirk, M., & Aouizerat, B. E. (2013). Current and emerging technology approaches in genomics. *Journal of Nursing Scholarship*, *45*, 5–14.

National Human Genome Research Institute. (2016). *All about the Human Genome Project*. Retrieved February 5, 2016, from http://www.genome.gov/10001772

National Institute of Nursing Research. (2016). *Implementing NINR's strategic plan: Key themes*. Retrieved February 5, 2016, from https://www.ninr.nih.gov/aboutninr/keythemes

Rigden, D. J., Fernandez-Suarez, X. M., & Galperin, M. Y. (2016). The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. *Nucleic Acids Research*, *44*, D1–D6.

Tully, L. A., & Grady, P. A. (2015). A path forward for genomic nursing research. *Research in Nursing & Health*, *38*, 177–179.