APPENDIX 1: Internal Manuscript Concept Sheet

eMERGE Network: Manuscript Concept Sheet

Reference Number *(to be assigned by CC)*	NT480
Submission Date	06.28.2023
Project Title	Clinical and psychosocial impacts of sex- and gender-related data collection and usage in genetics research: A qualitative analysis within a large genomics network
Tentative Lead Investigator *(first author)*	Makenna Martin, Harris Bland (co-first authors)
Tentative Lead Investigator Email *(first author)*	Makenna.martin@vanderbilt.edu; harris.bland@vumc.org
Tentative Senior Author *(last author)*	Kathleen F. Mittendorf (corresponding author: kf.mittendorf@vumc.org)
All Other Authors	Stephanie A. Kraft, Gillian Hooker, Gail Jarvik, Georgia L. Wiesner, Luke Rasmussen, Emma Perez, Katherine Anderson, Maureen Smith, Dan Roden, Digna Velez Edwards, additional interested eMERGE authors
Sites Participating	VUMC, MCG, NW, UW
Background / Significance	It is critical to collect gender identity data to ensure affirming and equitable genetics care for transgender, gender diverse, and sex diverse (TGSD) patients. Further, a person's phenotypic sex (expression of primary, secondary, and endocrinological sex characteristics), chromosomal sex, and organ/tissue inventory can directly impact crucial aspects of testing and care recommendations. However, TGSD individuals can experience psychosocial and clinical harms if gender/sex data collection and related care are not culturally competent and accurate. Realistic fears about such harms could deter TGSD individuals from participating in clinical research, and enrolled TGSD participants may experience harms from participation if data models and study processes do not appropriately account for these variables. Automated processes are increasingly used to streamline care and reduce health system burden, yet these processes increase the risk of misgendering participants or providing clinically inappropriate recommendations if these data are not accurate. The Electronic Medical Records and Genomics (eMERGE) Network is collecting sex/gender data at two points with different question/response options, and also has access to electronic health record data, offering an opportunity to evaluate the impact of a large electronic health record (EHR)-integrated genomics research study on TGSD individuals. The Network is evaluating the influence of a novel EHR-integrated integrated genome-informed risk assessment (GIRA) report on clinical care in 25,000 patients across 10 sites. The GIRA relies on data in the sex field to automate inclusion of recommendations on the report for certain conditions,

	and as a result has the potential to negatively impact TGSD participants. Further, the family history-based risk assessment, which supplies a component of risk information included in the GIRA, utilizes a separate sex field with a different question/response structure, creating the potential for misaligned data elements. In a supplement application to the Vanderbilt Genome-Electronic Records (VGER) Project, which supports the Vanderbilt University Medical Center site of the eMERGE Network, we poposed to use in-depth semi-structured qualitative interviews to investigate the impact of the eMERGE data collection, data usage, and processes on the validity of the GIRA for TGSD individuals and on potential psychosocial and clinical harms from the study on this population. We will use information gleaned in these interviews to develop a best practice guide for genomics research involving TGSD populations and propose a model for data collection that can support such research while remaining culturally competent and accurate. This model will be included in the FHIR specification for the GIRA. This supplement directly supports the overall project goals to validate and improve upon the GIRA. This concept sheet will support the qualitative analysis delineated by Aim 1 of the proposed eMERGE SGM supplement. A separate concept sheet will support the development of the proposed data model and FHIR specification.
Outline of Project	This concept sheet will cover Aim 1 proposed in the eMERGE supplement, with supplementary data regarding the MeTree elements collected from interviews with FOREST participants Aim 1. Utilize qualitative interviews to understand the impact of data collection, data usage, automated processes, and EHR integration on TGSD participants in eMERGE. 1a. Evaluate the impact of participant interpretations and reactions to the sex and gender identity questions in the eMERGE data model on data quality/validity and the validity of the GIRA. 1b. Identify potential individual- and group-level harms to TGSD populations from the eMERGE model.
Desired Data - Common Variables* *(Available from the CC)*	X Demographics ☐ICD9/10 codes ☐CPT codes ☐Phecodes ☐BMI ☐Common Variable Labs ☐Common Variable Meds ☐ Geocoding 2015 ACS variables ☐Other: Case/Control status
Other Desired Data *(Available from participating sites)*	*Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*

	legal sex, assigned sex at birth, and gender identity as recorded in the EHR, if this data is present in EHR.
Desired Genetic Data	☐eMERGE I-III Merged set (HRC imputed, GWAS) ☐eMERGE PGx/PGRNseq data set ☐eMERGEseq data set (Phase III) ☐eMERGE Whole Genome sequencing data set ☐eMERGE Exome chip data set ☐eMERGE Whole Exome sequencing data set XOther (not listed above): GIRA risk results (component-by-component) which may be relevant to or changed by information found in interview. No full genomic data sets are required.
Does project pertain to an existing eMERGE Phenotype?	☐Yes, if so please list XNo
Planned Statistical Analyses	N/A, qualitative only
Ethical Considerations	Scheduling and Interviewers will need to be approved to contact individuals who are not necessarily located at their site, if other sites agree to supply participants. Trans participant data is particularly sensitive, and qualitative transcripts may need to be additionally redacted beyond typical considerations prior to uploading for analysis. Trans participant data is particularly sensitive, and will need to have special reporting considerations when present in small numbers Trans participant data will be stored in limited access location, and all row-level data protected with password encryption. We will follow the data and safety monitoring plan we developed for a partner study evaluating the validity of the MeTree application for FOREST participants.
Target Journal	
Milestones *(This section should include the key dates for completion of project, including approval, project*	IRB submission: Mid-summer Interview Start goal: August 2023 Interview completion goal: December 2023 Interview Analysis completion goal: February 2024 Full draft: May 2024 Submission: Summer 2024

duration, draft completion, and submission.)	

******** *This section should include the timeline for completion of project, including: approval, project duration, first and second draft of the paper and submission.*

APPENDIX 4: Acknowledgement Text

eMERGE Network (Phase IV)

This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG011172 (Cincinnati Children’s Hospital Medical Center); U01HG011175 (Children’s Hospital of Philadelphia); U01HG008680 (Columbia University); U01HG011176 (Icahn School of Medicine at Mount Sinai); U01HG008685 (Mass General Brigham); U01HG006379 (Mayo Clinic); U01HG011169 (Northwestern University); U01HG011167 (University of Alabama at Birmingham); U01HG008657 (University of Washington); U01HG011181 (Vanderbilt University Medical Center); U01HG011166 (Vanderbilt University Medical Center serving as the Coordinating Center)