UNC
SCHOOL OF MEDICINE
DEPARTMENT OF GENETICS

A: COVER LETTER

Kellie B. Kelm, Ph.D
Division of Chemistry and Toxicology Devices
Office of In Vitro Diagnostics and Radiological Health (OIR)
Food and Drug Administration
10903 New Hampshire Avenue
WO66, Room 5648
Silver Spring, MD 20993-0002

Re: Pre-submission IDE Review Request

Device Name: 1-U19-HD077632-01 (Powell and Berg co-PIs) 9/5/2013 - 8/31/2018

Funding by NICHD and NHGRI

Dear Dr. Kelm:

Please find attached our pre-submission enquiry (2 hard copies and 1 e-copy) regarding the potential need for an IDE in order to carry out our above referenced NICHD/NHGRI-funded U19 project entitled "NC NEXUS," which examines the clinical application of exome sequencing for newborn screening.

The e-copy is an exact duplicate of the hard copies.

As you know, many leading voices in the field of human genome research have proposed that eventually every newborn will have his or her genome sequenced at birth, and that this information would provide guidance regarding their medical care throughout life. With that being said, such a scenario poses enormous technical, practical, ethical, and social challenges. It is with these challenges in mind that the NIH set out to fund several highly integrated projects to explore these issues in order to inform the use of genome-scale sequencing in newborns (http://grants.nih.gov/grants/guide/rfa-files/RFA-HD-13-010.html).

We at UNC approached our proposal somewhat skeptical of the general premise, but eager to participate in the research that will ultimately guide this important area of genomic medicine. ***We do not have any intent to commercialize a device or platform for genomic newborn screening. We are not planning to market this service commercially or to develop it as a companion diagnostic for any medication.*** As an academic research group, we are interested in studying the technical possibilities and limitations of next-generation sequencing technology, evaluating the psychosocial impact of decision-making on parents, and ultimately contributing to a body of knowledge that will help the field grapple with implementation of genomic screening in newborns.

The attached document provides a detailed description of the in vitro "device" (exome sequencing with targeted informatics analysis and CLIA Sanger confirmation) and the proposed intended use in our research protocol. We have considered carefully the potential risks involved in such research, both the psychosocial risks involved in this research in newborns and the technical risks related to device failure.

- Concerns about the psychosocial harms of returning information of a potentially sensitive nature (for example, the presence of genetic variants consistent with the presence of a Mendelian disorder) are largely a human research subject protection issue that we feel can be adequately addressed by providing skilled genetic counseling regarding any such results. These aspects of the study will furthermore be subject to appropriate oversight by the UNC Institutional Review Board.

- Concerns regarding technical device failure fall into two possible categories: 1) false positives introduced by next-generation sequencing; and 2) false negatives of the exome sequencing pipeline. The first of these concerns is mitigated by confirmation using orthogonal methods in the CLIA-certified MGL. The second concern is offset by the fact that every participant will have had standard of care newborn screening, and we will not be relying on the exome sequencing platform to provide any information that would not otherwise be clinically available.

Together, we feel that the proposed research is scientifically and ethically sound, and furthermore that pilot investigations such as these are a necessary prerequisite to more comprehensive assessments of the utility of next-generation sequencing for newborn screening.

We are happy to answer any additional questions that the FDA has and look forward to your feedback. Our specific questions are: What level of risk is involved in the proposed study? Will our proposed study require an IDE? What modifications of the protocol are recommended by the FDA? During the course of the study, what changes to the protocol or IRB would require additional review by the FDA?

Thank you for your thoughtful and expeditious review of this pre-submission IDE request. In addition to Drs. Powell and Berg, the following individuals may act with regard to this request:

Laura V. Milko, PhD
5100B Genetic Medicine Building, CB 7264
The University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7265
Phone: 919-843-2878
Email: laura_milko@med.unc.edu

Sincerely,

Cynthia M. Powell, MD
Professor of Pediatrics and Genetics
PI, NC NEXUS

Jonathan S. Berg
Assistant Professor of Genetics
PI, NC NEXUS

Attachment: CDRH premarket review submission cover sheet

B. TABLE OF CONTENTS

C. Overview of the Device:

Each of the steps involved in the NCNEXUS project can be envisioned as an element of a workflow: sample collection, laboratory methods (DNA extraction, exome library preparation, massively parallel sequencing), bioinformatics pipeline (initial informatics analysis and variant annotation), clinical interpretation of exome sequence variants (screening and indication-based analysis), variant confirmation in the CLIA-certified Molecular Genetics Laboratory, and finally, return of results. ***The "device" thus consists of exome sequencing (ES) with targeted informatics analysis and CLIA Sanger confirmation.***

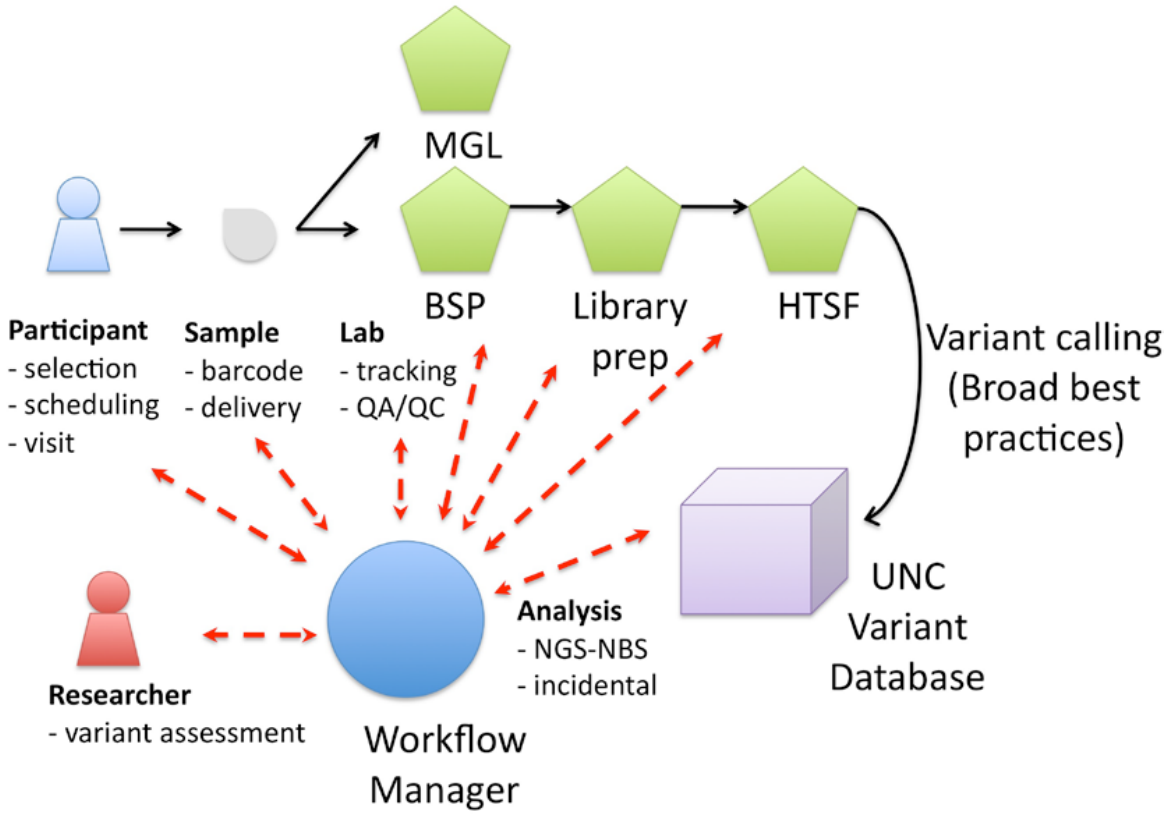
We will utilize a robust infrastructure for workflow management, sequencing library preparation, bioinformatics pipelines, data analysis and interpretation (see figure 1). This infrastructure is completely operational, highly flexible, and fully integrated with external databases and with other systems at UNC, including the Biospecimen Processing (BSP) Facility, the High Throughput Sequencing Facility (HTSF), and the UNC Hospitals CLIA-certified Molecular Genetics Laboratory (MGL).

1. Sample collection:

Upon enrollment, DNA will be collected non-invasively by swabbing the inside of subjects' mouths using Oragen DISCOVER (OGR-250) sample kits (DNA Genotek, Ontario, Canada) with CS-1 accessory sponges for assisted collection. Duplicate samples will be obtained, with one sent to the BSP and the other to the MGL. The BSP barcodes all samples and uses highly automated procedures (see https://genome.unc.edu/bsp/). The BSP has processed approximately 100,000 samples and generated almost 1 million aliquots for 76 different projects in the last 4 years. The MGL routinely isolates DNA from thousands of

samples per year for clinical genetic testing using automated methods validated for quality, including cheek swab samples.

Figure 1: NC NEXUS workflow schematic



Schematic representation of the NC NEXUS workflow management system, which coordinates elements such as visit scheduling, sample acquisition and labeling, laboratory steps including DNA purification and storage in the biospecimen processing facility (BSP), exome library preparation, massively parallel sequencing in the high-throughput sequencing facility (HTSF), sequence analysis and variant calling pipelines, storage and annotation of variants, and analytic algorithms.

The strategy of obtaining duplicate samples at the time of the enrollment encounter ensures rigorous quality control by identity screening using an efficient and cost effective panel of eight common SNPs that are genotyped in the MGL and compared to exome sequence variant calls at those positions. This approach has been employed in an ongoing research study

that uses the same pipeline, with 100% confirmation in ~200 samples, indicating that no sample swaps occurred during the exome sequencing process.

Duplicate samples also streamline the process of variant confirmation. If there is a discrepancy between the research exome sequence data and the Sanger sequencing data generated in the MGL, testing will be repeated in both labs using the independent samples. If DNA isolated by the BSP and MGL appears to have come from different subjects, a new sample will be obtained, subjected to repeat testing, and the source of error determined. Based on our experience in an ongoing project, we do not expect problems with sample misidentification, and duplication of cheek swab samples in the BSP and MGL will provide additional confirmation of accuracy and patient identity for results that are reported clinically.

2. Laboratory methods:

Our research group has successfully generated 200 high quality exome sequencing libraries for use in other studies at UNC. All protocols are followed rigorously to ensure a high level of reproducibility between samples.

i. DNA extraction: DNA will be extracted from salivary leukocytes and buccal cells shed from the inner lining of the mouth using a QIAamp® DNA Blood Mini Kit and supplementary protocol from the manufacturer (Qiagen, Valencia, CA).

ii. Exome library preparation: Our current protocol for library preparation utilizes the SureSelect XT Human All Exon V5 library (Agilent Technologies Inc. Santa Clara, CA). We have established methods for manual library preparation or automated protocols employing a 96-well format on a Caliper instrument programmed for use with Agilent protocols. Enriched libraries are tested for QC/QA for size distribution and

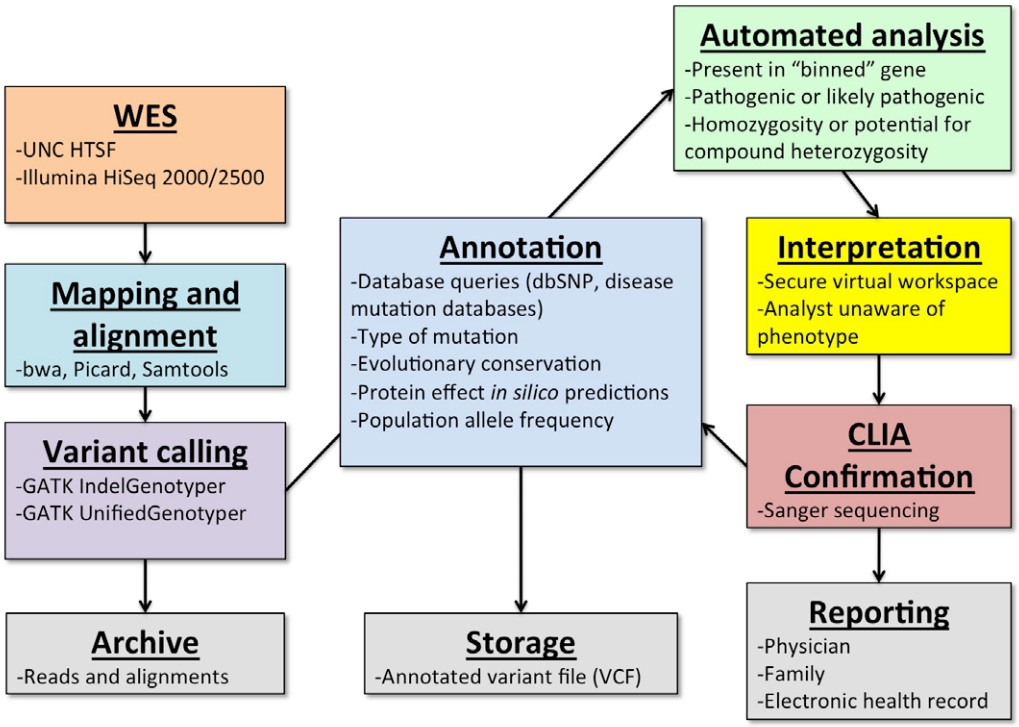
concentration using an Agilent Bioanalyzer 2100. Index barcodes are used in order to pool samples (currently four per pool).

iii. Massively parallel sequencing: Pools of samples will be subjected to massively parallel sequencing using either Illumina HiSeq 2000 or Illumina HiSeq 2500 sequencers that are housed, maintained, and operated by the UNC HTSF, which has provided high-quality raw sequence for several large sequencing initiatives at UNC. Currently, the HiSeq platform can produce more than enough sequence to generate 50-100x average coverage exome data when pooling four samples per lane. The HTSF operates under stringent quality control (QC) conditions: (1) DNA/RNA concentration is estimated based on fluorescent detection, (2) library quality is verified using the LabChip LX automated electrophoresis system (Caliper), providing information related to size of the inserts and level of contamination, and (3) analysis of sequencing data (e.g. sequence coverage, presence of adapter sequence, rRNA gene contamination).

3. Bioinformatics pipeline:

The Renaissance Computing Institute (www.renci.org) has been integrally involved in the development of an integrated pipeline for variant calling and analysis. Through a combination of existing and adapted computing tools coupled with traditional analysis tools, the bioinformatics pipelines are able to: (1) perform large scale computations including alignment and variant calling, (2) coordinate a pipeline of such calculations, (3) store reads, assemblies, variants, and annotations, (4) provide data sets to researchers, and (5) provide for efficient query of a large variant database. The RENCI team has built an infrastructure that integrates the technologies necessary to achieve these goals (see figure 2).

Figure 2: NC NEXUS bioinformatics pipeline



Schematic representation of the bioinformatics pipeline from raw data generation, to variant calling, annotation, variant filtering, human curation, and CLIA confirmation.

i. Initial informatics analysis (mapping, alignment, variant calling): The early bioinformatics steps required to generate sample-specific reads from multiplexed flow cells are performed using Casava. The resulting fastq files are then further processed using bwa to align reads to the current reference sequence. In addition to its considerable performance characteristics, bwa operates on paired-end reads, performs gapped alignments, and creates output in SAM format. Resulting SAM files of aligned reads are sorted, indexed, and converted to binary BAM files using Picard and Samtools. Post-alignment optimization, including PCR duplicate removal, realignment of reads, and quality score recalibration are performed using The Genome Analysis Toolkit (GATK). Indels are called from BAM files using the GATK IndelGenotyperV2 and single nucleotide polymorphisms (SNPs) are called using the GATK Unified Genotyper.

ii. Variant annotation: VCF files are annotated using a variant database, developed by RENCI, that calculates and stores annotations about all known and newly observed variants including those generated by several federated projects at UNC and external data such as the 1000 Genomes Project and NHLBI GO Exome project. The database utilizes built-in scripts to perform automatic updating of information from external sources and applies annotations to novel variants, with information such as transcript location, whether the variant affects a splice site, and type of mutation (e.g. missense, nonsense, or indel). These scripts will automatically import and archive new genome builds or reference transcript sets, and translate all data to the new reference system. The database can retrieve variant information based on any previously used reference sequence and integrate summary incidence data from sources that used different

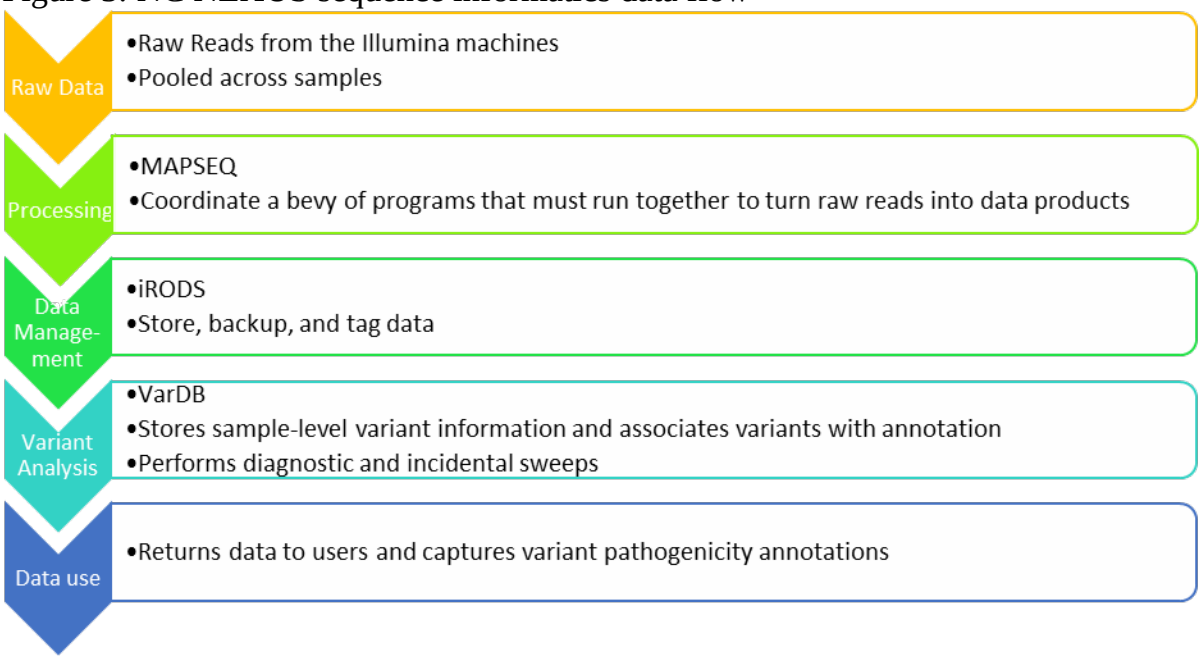
builds. An archival version of all data sources are kept such that it is possible to reconstruct a view of the data as it existed at any point in the past. Functional annotation of exome sequence variants leverages diverse types of information, including dbSNP identifier, occurrence in the Human Gene Mutation Database or other disease-specific mutation databases, frequency in control populations, and other annotations related to gene structure and protein effects of the variant. Further analysis, including protein structure information, sequence conservation, motif conservation, or other context-specific predictors are collected and calculated.

Quality metrics are captured at all stages of processing to determine if outputs can be used for analysis. Metrics include checks on input file correctness, distributions of nucleotide and quality scores, percent of reads aligned, read gap distributions, percent of reads with pairs, metrics on coverage across the genome and from targeted regions, and metrics from GATK on called variants.

Reads and alignment files in fastq and BAM format will reside at RENCI in a combination of network-attached storage (Multinode BlueArc Titan 3200 NFS/CIFS cluster, connected via 40Gbit Ethernet) and tape archive (Quantum Scalar i2000). These files will originally be stored to the NAS disks; once the raw data are processed through variant calling, iRODS will manage transfer to and from tape archive. The iRODS software will also be used to automate the flow of data from site to site, allow different sites to be treated transparently as a single data grid, and index read and alignment files, allowing researchers to locate particular reads quickly from within a massive set of data. Based on preliminary work with exome sequence data and storing the data from the 1000 Genomes Project we estimate we will require ~10 Gb of storage space

per sample to store raw reads in gzipped fastq format and mapped reads in BAM format. From this, we estimate that ~20Tb of storage space will be required for consensus sequences, variants and associated per read quality scores and raw and mapped reads in the map reduced queryable data store and database (see figure 3).

Figure 3: NC NEXUS sequence informatics data flow



4. Clinical Interpretation of Genomic Variants:

From an informatics standpoint, potentially clinically relevant variants will be selected using a series of automated filters and computational prioritization parameters that are applied to the individual's variant data. When one considers the requirements of a public health oriented newborn screening program, the informatics approach must be highly automated and efficient, producing actionable results with minimal human input. These analytic processes must provide results with high sensitivity while also minimizing false positive results (high specificity).

In the case of sequence variant data, we consider the analytic performance of the assay separately from the clinical performance of the assay. Analytic performance can be calculated via comparison with a different "gold standard" analytic method, namely Sanger sequencing. It will be straightforward to determine the analytic specificity of exome sequencing, since we will be confirming many called variants with Sanger sequencing. We will thus be able to detect false positive variant calls from the exome sequencing pipeline. Analytic sensitivity will be calculated by comparison to known clinical sequencing results of in patients with known genetic disorders (see Section D for details regarding the patient cohorts). Since it would be impossible from a practical standpoint for a human to review every variant, the clinical sensitivity and specificity depend on the molecular analysis processes used to select variants for further human review. Clinical sensitivity can be calculated by determining whether informatics analysis of the exome sequence data correctly identifies participants affected with a genetic disorder as having genetic variants that would predict the presence of that disorder. Clinical specificity is more challenging to calculate because although the variant data might predict the

presence or future development of a genetic disorder, there may not be a confirmatory test (such as enzyme testing) available. In these cases, longitudinal clinical follow-up will be required to determine the clinical specificity of the exome sequencing results.

In the NC NEXUS study, molecular analyses will generally fall into two categories that will use distinct computational algorithms:

i. Screening analysis mode: The first category of analysis is a "screening" mode, in which the prior probability that a given individual will have a rare Mendelian disorder is exceedingly low, typically much less than 1/1000. In this setting, the status of the individual will be unknown but presumed to be unaffected, as is the case for most healthy newborns, and we will use computational algorithms to select a small subset of variants contained within certain categories of genomic findings (described in section D.1 below) for human review and potential reporting. See section F for details on the development of these algorithms. In this setting, the analytic team will be blinded to the phenotypic status of the subject so as to produce unbiased reports, as would be the case in a "real world" NBS context.

ii. Indication-based analysis mode: The second category of analysis is an "indication-based" mode, in which there may be abnormal results on the traditional newborn screening assays (e.g. biochemical screening, hearing screening, etc.) or the postnatal development of phenotypic manifestations that might prompt reanalysis of exome sequence data (e.g. developmental delay, seizures, cancer). In this category, the prior probability of a given genetic disorder is higher and merits a more in-depth scrutiny of variant data. We will facilitate these analyses by filtering exome sequence data to select

variants within pre-developed gene lists appropriate for the clinical indication and computationally prioritizing variants for human review.

The resulting suspicious variants (if any) will be presented to the molecular analyst by the workflow engine, along with associated annotations and external links (e.g. OMIM, genome browsers to visualize the position of the variant, alignments of individual read data) that may be required to adjudicate the pathogenicity of the variant. The project team will conduct "molecular sign-out meetings" at which the results will be discussed and triaged for follow up in the MGL. Members of the CLIA-certified MGL will participate in the sign-out meeting to assist in determining which variants will be confirmed and reported.

5. Confirmation and reporting of exome sequence variants:

All sequence variants reported to patients will first be confirmed in the MGL, CAP-accredited and CLIA-certified for high complexity testing with a broad menu of clinical molecular diagnostic tests. The MGL is equipped with state-of-the art equipment for molecular diagnostic testing, including three ABI 3030*xl* genomic analyzers for Sanger sequencing analysis, Qiagen Pyromark MD for pyrosequencing analysis, and an Affymetrix GeneChip system for expression, copy number variation, and genotyping. The MGL currently performs multiple clinical assays utilizing Sanger sequencing, including custom sequencing to confirm the presence of genetic variants identified in a research setting or for testing of family members, and thus has considerable experience with design, analysis and interpretation of clinical Sanger sequencing. For each novel or rare variant, primers flanking the variant of interest will be designed with an M13 primer tag and subjected to bidirectional Sanger cycle sequencing.

In most cases, it is anticipated that Sanger sequencing will be appropriate for the confirmation of rare variants. However, it is possible that WES may identify mutations for which clinical testing is currently available but for which Sanger sequencing is not ideal. Because the diagnostic implications of these results are significant, in these rare cases the appropriate gold standard molecular diagnostic test will be performed. The MGL will issue a clinical report that details the confirmation and interpretation of any positive findings from the exome sequencing process. A separate "research report" will also be issued that describes the aggregate results of the exome sequencing process including coverage details and the total numbers of variants identified in different categories, but will not include any specific variant details.

D. Proposed Intended Use/Indications for Use

The NC NEXUS study will evaluate the use of exome sequencing as a potential means to augment newborn screening. The main technical outcome will be to examine the sensitivity and specificity of this technology in detecting conditions that are currently screened for in newborns. Another technical outcome will be to examine the capacity of exome sequencing to detect other conditions that would be beneficial to identify at an early age in children but for which there is currently no available diagnostic method. In addition to the examination of technical outcomes, the NC NEXUS project includes a highly integrated set of research aims that will address the ethical/legal/social implications (ELSI) aspects of exome sequencing in newborns.

1. Categorical subdivision of genomic findings:

One of the key questions being investigated in the NC NEXUS project is how best to divide the broad and heterogeneous range of genomic variants into categories that allow parents to make well-informed decisions about a) whether or not to pursue exome sequencing for their newborn; and b) what types of genomic information they are interested in learning. To accomplish this, we propose to use a method for "binning" the genome into categories based on clinical validity and clinical actionability.

The primary goal of screening in healthy newborns is to identify the existence of a condition that is preventable if detected before the onset of symptoms. This rationale can be broadened to include conditions in which early interventions may mitigate the symptoms, even if the disorder is not fully preventable. It has even been proposed that early recognition of a vast number of genetic disorders could reduce a patient's "diagnostic odyssey" and thus prove beneficial to families, even if ultimately untreatable. However, this expansion of the mission of newborn screening raises substantial ethical concerns and there is no consensus regarding how such information should be handled and very little evidence about the benefits or harms of providing non-medically actionable information to parents. In order to examine parental preferences and begin to explore the psychosocial impacts of providing such genomic information, we have designed a randomized trial in which certain categories of information will be offered to parents in the "experimental" arm and they will be asked to choose which information to learn. For this effort we have developed the following bin structure for the analysis and categorization of genomic findings:

- ***A. Medically actionable conditions with onset of symptoms and/or initiation of interventions in childhood***, representing the core "Next-Generation Sequencing Newborn Screen (NGS-NBS)". This category would include genes implicated in conditions that are currently screened for in extended newborn screens, including metabolic disorders, endocrine disorders, and hearing loss. In addition, we will include other medically actionable conditions that are not amenable to current screening methods but can be detected using genetic sequencing. These findings will represent the default set of results that would be returned with every exome sequence report.
- ***B. Medically actionable conditions with onset of symptoms and/or initiation of interventions in adulthood.*** This category would include genes implicated in conditions for which specific management strategies are recommended for individuals prior to onset of symptoms, and which would be expected to provide defined benefit in terms of reducing morbidity or mortality. The presence of such conditions may be of interest to some parents but not to others, and therefore parents in the experimental arm will be able to choose whether or not to learn of findings in this category. Parents in the control arm will not be given access to this information.
- ***C. Non-medically actionable childhood-onset health conditions.*** This category includes genes implicated in childhood-onset genetic disorders for which no specific preventive measure or treatment has been shown to mitigate morbidity. The presence of such conditions may be of interest to some parents but not to others, and therefore parents in the experimental arm will be able to choose whether or not to learn of findings in this category. Parents in the control arm will not be given access to this information.

- ***Non-medically actionable adult-onset conditions.*** This category includes genes implicated in adult-onset genetic disorders for which no specific preventive measure or treatment has been shown to mitigate morbidity. Although parents are given a great deal of latitude in determining how best to manage their child's healthcare, there are some conditions that many consider unethical to report in the context of a child. For example, it is widely accepted that predictive genetic testing for adult-onset conditions such as Huntington Disease or highly penetrant Alzheimer Disease that lack any prevention or treatment should not be performed in or reported to asymptomatic minors. This category of information will not be analyzed or offered to parents.
- ***Findings that provide information about reproductive risks.*** This category relates primarily to findings that have reproductive implications, such as carrier status for recessive disorders.

Other genomic variants that have no clear association with any genetic disorder (a category that encompasses the vast majority of the genome) will not be analyzed or reported to parents. This includes any loci that have not yet been directly linked to genetic disorders and are thus of no importance in a clinical context.

One of the research activities and deliverables planned for the NC NEXUS project is the development of a standardized procedure for categorizing genomic loci into these categories. See Section F for a description of the development of this metric.

2. Target populations:

In order to evaluate the performance characteristics (analytic sensitivity, analytic specificity, clinical sensitivity, and clinical specificity) of exome sequencing for newborn screening, we will include cohorts from the following patient populations:

- **Infants and children from 0-5 years with confirmed Phenylketonuria (PKU)** identified through newborn screening and followed in the UNC Pediatric Genetics and Metabolism Clinic. Currently the UNC Division of Pediatric Genetics and Metabolism follows 33 patients with PKU between 0-5 years of age and 5-7 new cases are ascertained each year. Younger patients are seen every 3 months and older ones at least once a year.
- **Infants and children from 0-5 years with confirmed medium chain acyl CoA-dehydrogenase deficiency (MCADD)** identified through newborn screening and followed in the UNC Pediatric Genetics and Metabolism Clinic. Currently the UNC Division of Pediatric Genetics and Metabolism follows 28 MCADD patients from 0-5 years of age and 5-7 new cases are ascertained each year.
- **Infants and children from 0-5 years with confirmed Cystic Fibrosis (CF)** identified through newborn screening, **those with positive screens but non-confirmatory follow-up testing (false positives)**, and a group with ***CFTR*-related metabolic syndrome (CRMS)** who have a positive CF newborn screening, fewer than two pathogenic CF mutations, and non-diagnostic sweat chloride levels. The UNC Pediatric Pulmonology Division sees 12-22 new patients with CF each year and 1-3 per year with CRMS and follow approximately 65 patients with CF from 0-5 years. UNC performs sweat chloride testing

on approximately 130 infants per year referred due to abnormal CF newborn screening, with more than 100 having negative testing.

- **Infants and children 0-5 years with confirmed hearing loss** identified through newborn hearing screening. UNC evaluates approximately 300 new patients each year for abnormal newborn hearing screening with sedated ABRs. Approximately 200 have confirmed sensorineural hearing loss or auditory neuropathy. These patients typically return several times each year for follow-up in the Pediatric Audiology and ENT Clinics. Approximately 1800 pediatric patients are followed at UNC. The standard clinical protocol at UNC for infants with confirmed sensorineural hearing loss is to obtain a newborn blood spot sample on their child from the State Newborn Screening Laboratory after the parent signs a release form, test for mutations in connexin 26 and 30 and perform CMV PCR.
- **Children with miscellaneous conditions that have been considered for recommended newborn screening panels but have not been adopted due to lack of available testing.** These patients are followed in Pediatric Genetics and Metabolism Clinic and Child Neurology Clinic and include such disorders as the mucopolysacharidoses, adrenoleukodystrophy, and Wilson disease.
- **Children with confirmed or suspected primary ciliary dyskinesia (PCD).** Approximately 20 patients from 0-5 years with known or suspected PCD are seen at UNC each year in the Pediatric Pulmonary Clinic.

- **Healthy newborns.** These infants will be ascertained prenatally by obtaining informed consent from their mothers who are followed in the North Carolina Women's Hospital's prenatal clinics and their fathers, if available.

3. Tissue sampling: We will use two buccal swabs to obtain DNA. One swab will be sent to the research lab for DNA extraction and exome sequencing. The other will be sent to the clinical lab for DNA extraction and storage. See section C.1 for details.

4. Frequency of use: We plan to obtain buccal swabs once from study participants unless DNA isolated by the BSP and MGL appears to have come from different subjects, in which case a new sample will be obtained and subjected to repeat testing.

5. Physiological use: Not applicable.

E: Previous Discussion or Submissions:

None.

F: Overview of Product Development

The device (exome sequencing with targeted informatics analysis and CLIA Sanger confirmation) is not intended for any type of prescription or over-the-counter use. It is not being developed for any type of commercial use. It is only being utilized in a research setting and all subjects/parents of subjects will be consented through a detailed and lengthy process of consent (informed decision making). Because of the nature of this pilot project, many aspects of the device will be under continuous development over the course of the research project.

1. Development of informatics pipelines and algorithms:

Since this is a research project, we anticipate that there will be periodic alterations of the sequencing informatics pipelines, including updated genomic reference sequence and

annotations, and algorithms for mapping, alignment, and variant calling. We also plan to examine algorithms for calling copy number variants. Thus, the informatics pipeline is expected to be under nearly continuous development. Similarly, a major research component of the NC NEXUS project is the development of algorithms that will be used to select variants for human review. As such, we will be systematically evaluating different combinations of algorithms in order to maximize the analytic sensitivity, analytic specificity, clinical sensitivity and clinical specificity.

As noted in section C.3, the workflow management system that has been established for use with the NC NEXUS project will include native version control, such that for any given analysis it will be possible to know the specific version of each component of the pipeline used to generate the data. This feature also gives us the capability to compare the variants selected by any given combination of algorithms to any other combination of algorithms.

2. Determination of categorical subdivisions

As described in section D.1, one of the major research activities associated with the NC NEXUS project is the development of a standardized method to segregate genetic conditions into categories for use in decisions about return of results. We will use a semi-quantitative metric to assess the medical actionability of gene-phenotype pairs (genetic disorders). This metric explicitly recognizes that medical actionability is not a binary state, but a continuum. We identified five core characteristics of medical actionability, with particular emphasis on the ramifications of finding a pathogenic variant in a person without signs or symptoms of the disease that would be predicted by the presence of the variant. These characteristics are reflected by the following questions:

1) Severity: “What is the nature of the threat to health for an individual carrying a deleterious allele in this gene?” (sudden death, possible death, serious morbidity, modest or no morbidity)

2) Likelihood: “What is the chance that a serious threat will materialize?” (> 50%, 5-49%, 1-5%, <1 %)

3) Efficacy: “How effective are interventions for preventing the harm?” (highly effective, modestly effective, minimally effective, ineffective)

4) Acceptability: “How acceptable are the interventions in terms of the burdens or risks placed on the individual?” (highly acceptable, moderately acceptable, minimally acceptable)

5) Knowledge base: “How much is known about the condition in order to score each category?” (substantial evidence, moderate evidence, minimal evidence, controversial or poor evidence)

All five criteria are given a score on a scale of 0-3. Scores for categories 1 and 2 are linked to the same outcome, either the most severe potential outcome or the earliest severe outcome that is typical for a given condition. Similarly, scores for categories 3 and 4 should reflect interventions targeted to the outcome described in categories 1 and 2.

In addition to the semi-quantitative “medical actionability” score, we will also determine the typical age of onset and the age at which medical interventions (if any) would be initiated. Together, these values can be used to define the four categories outlined in section D.1, by establishing a threshold for the actionability cut-off and a threshold for the age cut-off (see figure 4). We anticipate carrying out the scoring of different genetic disorders over the course

of the first year of the NC NEXUS project. In addition, as knowledge accrues about genetic disorders and their prevention and management, we will reassess the disorders and potentially reclassify disorders into a different category.

Figure 4: Categorization of genetic disorders using a semi-quantitative "binning" metric

A. Medically actionable conditions with onset of symptoms and/or initiation of interventions in childhood - High "actionability" score - Recommended initiation of interventions during childhood	***B. Medically actionable conditions with onset of symptoms and/or initiation of interventions in adulthood*** - High "actionability" score - Recommended initiation of interventions during adulthood
C. Non-medically actionable childhood-onset health conditions - Low "actionability" score - Childhood age of onset of symptoms	***D. Non-medically actionable adult-onset conditions*** - Low "actionability" score - Adult age of onset of symptoms

G: Specific Questions

- What level of risk is involved in the proposed study?
- Will our proposed study require an IDE?
- What modifications of the protocol are recommended by the FDA?
- During the course of the study, what changes to the protocol or IRB would require additional review by the FDA?

H: Mechanism for Feedback:

Our preferred method of feedback from the FDA is via teleconference involving the Joint Project Investigators, Cynthia Powell and Jonathan Berg, and the Project Manager, Laura Milko.

PIs:

Cynthia Powell - powellcm@med.unc.edu

Jonathan Berg - jonathan_berg@med.unc.edu

Project Manager:

Laura Milko - laura_milko@med.unc.edu