Scientific Goals

CZI Pediatric Networks grants have the following overarching scientific goals:

- Generate and share data that will contribute to and establish references for organs and tissues during age windows that will clarify childhood development and maximize utility for understanding pediatric disease;
- Represent diverse ancestries in data collected, with the goal of generating references that will promote understanding of disease in populations and communities with increased incidence of childhood disease;
- Build international collaborations that promote coordinated and cohesive progress and contribute to the Human Cell Atlas;
- Identify and address shared computational or infrastructural bottlenecks, to ensure that data generated as a part of this reference can be analyzed, incorporated, and shared as a part of a global reference;
- Collect and share biospecimens, including generating and disseminating solutions for challenges associated with ancestral diversity and vulnerable status of young donors, engagement with communities and families/patients, and storing tissue for subsequent single-cell analysis; and

https://chanzuckerberg.com/rfa/single-cell-pediatrics/

A Global Pediatric Cell Atlas of Nasal and Oral Mucosa

TEAM
Jose Ordovas-Montanes, PhD
Bruce Horwitz, MD, PhD
Lael Yonker, MD
Wanda Gonzalez, MD
Sarah Glover, DO
Partha Majumder, PhD
Senjuti Saha, PhD
Thushan de Silva, MBChB, PhD
Abdul Sesay, PhD
Alex Shalek, PhD
Corrine Sin Quee-Brown, MBBS

ABSTRACT
The nasopharyngeal and oral mucosa represent the initial sites of interaction with many environmental agents and microbes. Recent single-cell studies have revealed a rich diversity of epithelial and immune cell types and states within nasopharyngeal epithelium in diseases of global significance, including allergic inflammation and viral infection. Yet, beyond an accessible window into disease biology, minimally-invasive sampling of the nose and mouth in children represents a truly unique opportunity to characterize healthy mucosal epithelial and immune function worldwide. However, a comprehensive map of epithelial and immune system development across diverse ancestries and environments is lacking. To more broadly investigate the nasal and oral mucosa and understand how the normal variation present in healthy children maintains health or may inform disease, we have assembled an interdisciplinary team: unifying experts in 7 cities in 5 countries who are deeply invested in understanding the single-cell biology of the nasopharyngeal and oral mucosa in children living within our communities. Our plan aims to generate scientific and community engagement in all phases of our research to establish the foundation in Years 1 and 2 that will enable us to carefully and considerately analyze 80 pediatric participants at each site (560 total) across the age range from 1-month to 18-years of age in the next phase of our network (potential Year 3 and beyond). Our team will pilot and analyze single-cell data jointly with scientists from all locations, and share important lessons in global science, protocols and resultant data openly with the community. Ultimately, our global single-cell based characterization of the developing nasopharyngeal mucosa will reveal principles of epithelial and immune system development that will facilitate the equitable development of novel therapies for diseases of the aerodigestive tracts.

SCIENTIFIC GOALS OF THE PROJECT

Motivation: The nasopharyngeal and oral mucosa are the initial sites of interaction with many environmental agents and represent uniquely-accessible windows into **healthy,** rather than disease-adjacent, mucosal epithelial and immune function[1-5]. Outside of the neonatal period, respiratory disease is the leading cause of mortality for children under 5 worldwide, and environmental exposures including adolescent oral tobacco use may precipitate malignant transformation[6-9] (see **Project Details: Disease Relevance**). Healthy, developmental nasopharyngeal and oral mucosa reference data on children[10] from diverse locations and ancestries—reflecting global differences in susceptibility to communicable[11,12] and noncommunicable[13] diseases—is a critical unmet need**(Fig. 1)**. Recent single-cell and bulk transcriptomic studies by us and others have revealed unexpected complexity within nasopharyngeal/oral epithelium and associated immune cells suggesting the need for a deeper understanding of this tissue and its function in health and disease[14-22]**(Figs. 2, 3 and 4)**. We **hypothesize** that age, ancestry, and geographical location will strongly influence the development, differentiation, and function of mucosal epithelial and immune cell populations that regulate responses to viral, inflammatory, allergic, and toxic challenges. An unbiased and equitable single-cell RNA-

sequencing (scRNA-seq) characterization of the healthy developing nasopharyngeal and oral mucosa will reveal principles of epithelial and immune system development and health around the world that will facilitate the equitable development of novel therapies[23].

Revised Goals: We propose to establish our network and provide pilot data within a 2-year funding period. Our plan is rooted in scientific and community engagement as key enabling aspects to building a sustainable network to study the developing pediatric nasal mucosa across the globe. **(Fig. 1)**:

Goal-1) nucleate an international network of scientists dedicated to fostering a collaborative, equitable, and multi-disciplinary scientific and analytical environment;

Goal-2) recruit 140 healthy children aged 1mo-5yr, and acquire nasopharyngeal (for potential scRNA-seq) and buccal (for SNP-based demultiplexing) swabs, through engaging with interested families of diverse ancestry and socioeconomic development in 7 cities in 5 countries;

Goal-3) begin to establish the clinical, technical, and computational platforms to equitably generate an scRNA-seq global atlas of the nasopharynx from children (n=5-10 sequenced participants per site by end of Year 2) and the ~~oral cavity of adolescents (n=0)~~;

Goal-4) deeply analyze and annotate the cell types, subsets, and states present in the nasal and oral mucosa, and provide preliminary evidence for how the co-variates of ~~age~~, location, and ancestry influence cellular composition and phenotypes;

Goal-5) deploy/train scRNA-seq experimental methods, share data, and perform integrative analyses with other CZI Pediatrics and broader programs.

Goal-1) Our network is designed to ensure strong team science**(Fig. 1)**. Our team is composed of 11 co-PIs, welcoming new members to an ongoing collaborative core[15,23], with broad expertise including community-engaged researchers (n=7), pediatricians (n=4), single-cell biologists (n=4) and computational biologists (n=4). We have carefully integrated input from the extended scientific groups each co-PI represents to arrive at the current proposal; reinforcing our regular interactions. We will interact through bi-weekly zoom meetings amongst co-PIs, co-Is and trainees, with a rotating focus on community engagement, sample acquisition, single-cell capture, data analysis/sharing, and publications. We have established a shared Google Drive and github for sharing protocols and best practices, and a Slack channel where members from the network can interact in real time. We will deploy our established capacity[24,25] of training and enabling scRNA-seq capture globally to all sites during the scRNA-seq pilot phase (0-12 months). All resultant sequencing data and cell-by-gene matrices will be harmonized through data processing and normalization pipelines on a password-controlled Google Terra repository and a cellxgene virtual machine instance for scientists from all sites to jointly analyze scRNA-seq data.

Goal-2) All sites have demonstrated capacity for recruitment of pediatric participants and biobanking samples, including swab-based studies[8,9,13,26-29]. We will recruit healthy children and engage their families by leveraging expertise available at all sites in effective community engagement including discussion regarding the importance of medical research. We will capture self-reported ancestry (race/ethnicity), social determinants of health (e.g. food security), potential environmental exposures of significance (e.g. household smoking status), and history of allergic phenomena (food allergy, seasonal rhinitis, or mild intermittent asthma). Exclusion criteria will be standardized and include acute respiratory illness, current antibiotic usage, premature birth, persistent asthma, chronic lung disease, autoimmune disease, sickle cell anemia, immunodeficiency, and history of malignancy. Based on our recent and ongoing work profiling adult nasopharyngeal cells[15], we will sample and viably cryopreserve nasopharyngeal ~~and buccal swabs~~ from 20 children spanning 2 age-brackets (1 month-2 years(infants), 2-5 years(young child), ~~5-12 years(child) and 12-18 years(adolescent)~~) for a total of 20 participants and 20 potential scRNA-seq samples per site. We will also take one additional buccal swab for immediate RNA preservation and bulk RNA-seq. We have validated that our cryopreservation pipeline does not eliminate any cell types found in fresh healthy nasal swabs, and furthermore allows us to resolve disease-

specific biology with ~15 samples per group[15]. In Boston, MA, we will perform further extension of our fresh vs. frozen comparison (n=10) and share this as a resource with CZI Pediatric networks. This sampling strategy allows for single-cell processing and capture efficiencies by batching samples and demultiplexing based on natural genetic variation[30,31]. In the current initial stage of the study, wWe will profile nasopharyngeal swabs (n=5-to-10) ~~and oral mucosal (n=0 adolescent)~~ swabs from each site by scRNA-seq, with the capacity to scale based on our tiered **Tissue Resource**. Healthy participants representative of their communities (**Ancestral Diversity Table**) will be recruited during well-child visits, vaccination clinics, household visits, and/or emergency visits at 7 sites in 5 countries:

- **Boston, MA, USA**
- **Chelsea, MA, USA**
- **Dhaka, Bangladesh**
- **Jackson, MS, USA**
- **Kolkata, India**
- **Nassau, The Bahamas**
- **Serrekunda, The Gambia**

Goal-3) We strongly believe that the pursuit of scientific knowledge is a human right[32] and should be equitably distributed worldwide[23]. All participating sites have demonstrated expertise in conducting transcriptomic studies**(Figs. 2-5)** and/or have established a plan to enhance capacity. Scientists in each city will perform dissociation of cells[15] from nasopharyngeal swabs and massively-parallel scRNA-seq capture by either Seq-Well S^3 and 10X Genomics 3' V3.1. Due to the cost-effectiveness and rapid deployment of Seq-Well, which requires limited peripherals, we will enable local single-cell capture and library preparation capacity at all sites through the sharing of Seq-Well devices/reagents and on-site training, as demonstrated previously. For sites capable of running 10X, each site will select the most appropriate approach, given personnel/experimental constraints. These techniques are comparable in resolving matched cell types, subsets and states in direct comparison studies[24,33]. Seq-Well arrays will be loaded with individual participant samples. 10X Genomics channels will be loaded with three participant samples and SNP-based demultiplexing (jointly-collected bulk RNA-seq from additional oral swab) will be employed; as shown previously[30,31] and in ongoing projects in our laboratories**(Fig. 6)**. In order to ensure data interoperability, both within our network and with other CZI-funded projects, we will provide experimentally-driven integration anchors by performing matched Seq-Well and 10X profiling on previously collected reference nasal swabs at one site[33,34]**(Goal-5)**. Generating this reference data on two platforms is essential for CZI Networks seeking to build global references, as technologies like 10X are inaccessible to many scientists. We have already carefully considered optimal workflows at all sites accounting for current clinical and basic research workflows to arrive at the current budgeted workplan **(Budgets+LoS)**. To ensure rigorous data collection, all projects will have a 1 year pilot phase to fully vet the nasal swabs-to-single-cell capture and library generation/sequencing pipelines before selecting either the Seq-Well S^3 or 10X pipeline: fully adopting optimized methodology for the rest of the study. We will ensure that platform effects do not drive or confound biological data interpretation by statistically testing reproducible features[35]. Single-cells will be sequenced to a depth of 40,000 reads/cell on Illumina NextSeq or NovaSeq sequencers in Boston, Dhaka, Serrekunda, or Kolkata.

Goal-4) We will jointly analyze and annotate the cell types and states present in the nasopharyngeal/oral samples collected and sequenced at all sites. These analyses will be carried out in the Google Terra environment and will be led by computational biologists and experimentalists at all sites in order to help synergize data with collective expertise. Based on our previous studies, we anticipate that nasopharyngeal swabs will allow us to resolve the following major epithelial cell subsets: basal, secretory, goblet, ciliated, deuterosomal, ionocytes, enteroendocrine and squamous cells[15,17,18]. We will also capture the following

immune cell subsets: T cells, B cells, macrophages, dendritic cells and mast cells**(Figs. 2, 3 and 4)**. We anticipate buccal swabs will yield similar overarching cell types based on published data[36,37] and our own scRNA-seq data on oral cancer samples and flow cytometry on buccal swabs**(Fig. 5)**. We will analyze data to generate consensus markers, ~~as well as location- (n=80/site) and age-specific (n=140/bracket) marker gene lists,~~ for all cell subsets. We also anticipate based on the estimated scale of the dataset (0.5 to 1 million high-quality cells) generated that we will resolve novel cell subsets and states, along with differences in dynamic epithelial differentiation trajectories, upstream drivers, and predicted cell-cell interactions[15,38-40]. As ~30% of healthy children have a detectable asymptomatic viral infection when sampled**(Fig. 4)**, we will leverage our recent metatranscriptomic pipeline for scRNA-seq data to mine for "environmentally-present" viruses worldwide and compare subgroups based on this exposure category[41]. We will utilize covariates of age, gender, location, platform and ancestry for mixed-effects differential expression models to identify correlates with specific cell types and states and individual genes[42]. We will host quarterly hackathons in addition to regular lab-meeting style presentations (**Goal-1**)—focused on cell type/subset annotations and specific analytical techniques—with participation from all sites.

Goal-5) We will share data using our global health portal (Alexandria), CZI cellxgene, and HCA DCP, as we have done for other studies[15,17,22]. Through these methods, we will seek to leverage and grow interactions with other CZI Pediatric networks including novel ways of cross-referencing data through user-curated gene lists in other networks. We also propose a specific learning opportunity where we will run a workshop from barrier tissue cells-to-sequence in India open to CZI+HCA members to learn how to operate Seq-Well and 10X[23]. Furthermore, we will use this opportunity to generate the key reference data set (10x and Seq-Well anchor-gene data set from n=10 matched samples) that will help other studies using multiple methods rigorously compare data beyond computational integration. We will also develop graphical material and courses tailored to the families and participant age groups (middle school and older; **LoS**) at all sites to teach them about the principles of single-cell biology, and why understanding the nose and mouth is an important undertaking in the context of diseases that affect these tissues and the airways. Given the critical importance of language, we will disseminate material in the predominant local written language (**Community Engagement**).

Benefit from CZI Pediatric networks: We anticipate that CZI networks will contribute data from other mucosal sites including the intestine and lung. Developing methodology to compare findings between these tissues and the nasopharynx could reveal universal principles of mucosal development in children, as well as highlight important tissue-specific variation[10]. While we propose an ambitious cohort size for an scRNA-seq study, to answer questions at epidemiological scale, we will need to work together with other groups sampling these tissues to further scale numbers, further diversity, and address critical determinants of health such as social, climatic and environmental factors. This will require collective efforts across multiple networks profiling children in other locations.

Timeline for Year 1 to 2 to set up Year 3 and Future Years:

Year 1:

- Develop detailed community engagement plan at each site, full team review, and implementation.
- Validate sample collection process, storage, and resultant cell quality across all sites.
- Cross-platform training and comparison of 10X and Seq-Well single-cell capture and library preparation in India. This resource will facilitate data integration for this and other CZI-funded teams.
- Optimization of single-cell capture and library construction at each site. This will expand single-cell sequencing capacity in Jackson, Kolkata, Serrekunda, Dhaka, and Nassau.
- Recruit and biobank first 20 children at each site.
- Develop and implement data storage, sharing, and team analysis plan on pilot data.

Year 2:

- Recruit and biobank additional 60 children at each site (where capacity allows).
- Reduction in numbers: Perform single-cell capture, library preparation and sequencing for 5 to 10 participant samples as a pilot, per site, on chosen platform.
- Perform bulk RNA-seq to obtain reference transcriptomes and SNP-based data for demultiplexing.
- Pre-print posted to medRxiv and data to cellxgene describing cohort and sharing short-term analyses on 35 to 70 samples total.
- Interim dissemination of raw and processed data to other CZI-Pediatrics Networks to gain analytical power.
- Teaching in local middle and high schools.

Potential Year 3:

- Year 2 to 3 shift: Perform single-cell capture, library preparation and sequencing up to 25 participant samples per site
- Continued biobanking of samples at sites as capacity/community interest allows.
- Deep analysis and annotation of cell types, subsets and states present in the nasal and oral mucosa.
- Analysis of how co-variates of age, location and ancestry influence composition and state internally and with other CZI-Pediatrics Networks.
- Biological interpretations, manuscript preparations, data dissemination, and integration with data from other CZI-funded teams.

Plans for Year 4 and Beyond:

- Identify additional funding to support the continued working together for the network to biobank samples, continue to build infrastructure, continue community engagement, and continue joint analytical efforts.
- Identify additional funding to power single-cell capture, library preparation and sequencing for all 100 participant samples (80 nasopharyngeal, 20 buccal) per site for a total of 700 samples from 560 participants.

TOOLS AND RESOURCES

We have developed and shared via protocols.io a nasopharyngeal cryopreservation and scRNA-seq pipeline that is amenable for this and other studies[15]. We will continue to optimize this protocol through our work and share updates. Our team has pioneered the development and improvement of the cost-effective scRNA-seq Seq-Well platform which has been deployed in multiple countries worldwide[24,25]. Through this mechanism, we plan to enable Dhaka, Nassau, and Serrekunda to become full Seq-Well processing sites for this study, and to run a training session in India to enable future studies where other CZI programs with Asian hubs could join. Furthermore, we have deep expertise in studying the nasopharyngeal mucosa and engagement with the HCA community[22]. Highlights from our own studies and those of our colleagues using nasopharyngeal sampling include the reduction in epithelial cellular diversity and discovery of allergic inflammatory memory in chronic rhinosinusitis, novel epithelial cellular states in asthmatic individuals, the identification of SARS-CoV-2 viral-positive cells in individuals with COVID-19[14-22]. We will use this expertise to rapidly generate draft atlases of the pediatric nasopharyngeal mucosa. Furthermore, we will hope to leverage novel developments in cellxgene for automated cell type annotation, custom gene scoring, and ability to “gate” on specific cell types to ask targeted questions across our cohort and others.

TISSUE RESOURCES

Many current studies of mucosal barrier tissues in pediatric patients focus on disease, disease-adjacent, or symptomatic sampling rather than healthy tissue due to the ethical principle of avoiding unnecessary risk to healthy children. Further, advanced endoscopic procedures required to obtain biopsies of mucosal tissue are simply not performed on children in many regions of the world. Therefore, the non-invasive approach to obtain samples proposed here using nasopharyngeal and oral swabs is ideal for understanding mucosal **health** due to its low-risk and now common acceptance in society as a method for SARS-CoV-2 testing. Here we propose to prospectively collect nasopharyngeal/buccal swabs from 80 children across 4 age brackets at each of 7 sites. We have already biobanked swabs for ongoing work of diseased tissue at 4 sites, establishing feasibility. In this current budget, we propose to generate 100 single-cell sequenced samples in our initial budget amounting to 80 nasopharyngeal (where we have extensive preliminary data) and 20 buccal (exploratory adolescent) swabs at each site. However, due to the cryopreservation and batched-processing approach, we can tier these numbers accordingly (**Budget**), for example beginning with specific age brackets. Furthermore, it is quite likely based on the proposed success of our community engagement plans that we will recruit and biobank considerably more than 80 participants at each site over three years. In this event, further nasopharyngeal swabs may be available for orthogonal validation assays, or to share with other CZI Pediatric groups based on their expertise. All data we generate will be shared with the CZI and broader scientific community.

REFERENCES

1 Iwasaki, A., Foxman, E. F. & Molony, R. D. Early local immune defences in the respiratory tract. *Nat. Rev. Immunol.* **17**, 7-20, doi:10.1038/nri.2016.117 (2017).

2 Ordovas-Montanes, J., Beyaz, S., Rakoff-Nahoum, S. & Shalek, A. K. Distribution and storage of inflammatory memory in barrier tissues. *Nat. Rev. Immunol.*, doi:10.1038/s41577-019-0263-z (2020).

3 Gaffen, S. L. & Moutsopoulos, N. M. Regulation of host-microbe interactions at oral mucosal barriers by type 17 immunity. *Sci Immunol* **5**, doi:10.1126/sciimmunol.aau4594 (2020).

4 Moutsopoulos, N. M. & Konkel, J. E. Tissue-Specific Immunity at the Oral Mucosal Barrier. *Trends Immunol.* **39**, 276-287, doi:10.1016/j.it.2017.08.005 (2018).

5 Hewitt, R. J. & Lloyd, C. M. Regulation of immune responses by the airway epithelial cell landscape. *Nat. Rev. Immunol.*, doi:10.1038/s41577-020-00477-9 (2021).

6 Zar, H. J. & Ferkol, T. W. The global burden of respiratory disease-impact on child health. *Pediatr. Pulmonol.* **49**, 430-434, doi:10.1002/ppul.23030 (2014).

7 Mackenzie, G. A. *et al.* Respiratory syncytial, parainfluenza and influenza virus infection in young children with acute lower respiratory infection in rural Gambia. *Sci. Rep.* **9**, 17965, doi:10.1038/s41598-019-54059-4 (2019).

8 Saha, S. *et al.* The Direct and Indirect Impact of SARS-CoV-2 Infections on Neonates: A Series of 26 Cases in Bangladesh. *Pediatr. Infect. Dis. J.* **39**, e398-e405, doi:10.1097/INF.0000000000002921 (2020).

9 Yonker, L. M. *et al.* Pediatric Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): Clinical Presentation, Infectivity, and Immune Responses. *J. Pediatr.* **227**, 45-52 e45, doi:10.1016/j.jpeds.2020.08.037 (2020).

10 Taylor, D. M. *et al.* The Pediatric Cell Atlas: Defining the Growth Phase of Human Development at Single-Cell Resolution. *Dev. Cell* **49**, 10-29, doi:10.1016/j.devcel.2019.03.001 (2019).

11 Bhattacharyya, C. *et al.* SARS-CoV-2 mutation 614G creates an elastase cleavage site enhancing its spread in high AAT-deficient regions. *Infect. Genet. Evol.* **90**, 104760, doi:10.1016/j.meegid.2021.104760 (2021).

12 Biswas, N. K. & Majumder, P. P. Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type. *Indian J. Med. Res.* **151**, 450-458, doi:10.4103/ijmr.IJMR_1125_20 (2020).

13 Chatterjee, S. & Majumder, P. P. Kalyani cohort - the first platform in Eastern India for longitudinal studies on health and disease parameters in peri-urban setting. *Glob Health Epidemiol Genom* **2**, e2, doi:10.1017/gheg.2016.19 (2017).

14 Dwyer, D. F. *et al.* Human airway mast cells proliferate and acquire distinct inflammation-driven phenotypes during type 2 inflammation. *Science Immunology* **6**, eabb7221, doi:10.1126/sciimmunol.abb7221 (2021).

15 Ziegler, C. G. K. *et al.* Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. *bioRxiv*, doi:10.1101/2021.02.20.431155 (2021).

16 Cao, Y. *et al.* Single-cell analysis of upper airway cells reveals host-viral dynamics in influenza infected adults. *bioRxiv*, 2020.2004.2015.042978, doi:10.1101/2020.04.15.042978 (2020).

17 Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **181**, 1016-1035 e1019, doi:10.1016/j.cell.2020.04.035 (2020).

18 Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649-654, doi:10.1038/s41586-018-0449-8 (2018).

19 Ruiz Garcia, S. *et al.* Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* **146**, doi:10.1242/dev.177428 (2019).
20 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324, doi:10.1038/s41586-018-0393-7 (2018).
21 Vieira Braga, F. A. *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153-1163, doi:10.1038/s41591-019-0468-5 (2019).
22 Ballestar, E. *et al.* Single cell profiling of COVID-19 patients: an international data resource from multiple tissues. *medRxiv*, 2020.2011.2020.20227355, doi:10.1101/2020.11.20.20227355 (2020).
23 Majumder, P. P., Mhlanga, M. M. & Shalek, A. K. The Human Cell Atlas and equity: lessons learned. *Nat. Med.* **26**, 1509-1511, doi:10.1038/s41591-020-1100-4 (2020).
24 Hughes, T. K. *et al.* Second-Strand Synthesis-Based Massively Parallel scRNA-Seq Reveals Cellular States and Molecular Features of Human Inflammatory Skin Pathologies. *Immunity* **53**, 878-894 e877, doi:10.1016/j.immuni.2020.09.015 (2020).
25 Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398, doi:10.1038/nmeth.4179 (2017).
26 Lindsey, B. B. *et al.* Effect of a Russian-backbone live-attenuated influenza vaccine with an updated pandemic H1N1 strain on shedding and immunogenicity among children in The Gambia: an open-label, observational, phase 4 study. *Lancet Respir Med* **7**, 665-676, doi:10.1016/S2213-2600(19)30086-4 (2019).
27 Jarju, S. *et al.* Viral Etiology, Clinical Features and Antibiotic Use in Children <5 Years of Age in the Gambia Presenting With Influenza-like Illness. *Pediatr. Infect. Dis. J.* **39**, 925-930, doi:10.1097/INF.0000000000002761 (2020).
28 Cosgrove, P. R., Redhu, N. S., Tang, Y., Monuteaux, M. C. & Horwitz, B. H. Characterizing T cell subsets in the nasal mucosa of children with acute respiratory symptoms. *Pediatr. Res.*, doi:10.1038/s41390-021-01364-2 (2021).
29 Spence, D. *et al.* Advancing cancer care and prevention in the Caribbean: a survey of strategies for the region. *Lancet Oncol.* **20**, e522-e534, doi:10.1016/S1470-2045(19)30516-9 (2019).
30 Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89-94, doi:10.1038/nbt.4042 (2018).
31 van der Wijst, M. G. P. *et al.* Longitudinal single-cell epitope and RNA-sequencing reveals the immunological impact of type 1 interferon autoantibodies in critical COVID-19. *bioRxiv*, 2021.2003.2009.434529, doi:10.1101/2021.03.09.434529 (2021).
32 Wyndham, J. M. & Vitullo, M. W. Define the human right to science. *Science* **362**, 975, doi:10.1126/science.aaw1467 (2018).
33 Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737-746, doi:10.1038/s41587-020-0465-8 (2020).
34 Mereu, E. *et al.* Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* **38**, 747-755, doi:10.1038/s41587-020-0469-4 (2020).
35 Martin-Gayo, E. *et al.* A Reproducibility-Based Computational Framework Identifies an Inducible, Enhanced Antiviral State in Dendritic Cells from HIV-1 Elite Controllers. *Genome Biol.* **19**, 10, doi:10.1186/s13059-017-1385-x (2018).
36 Eipel, M. *et al.* Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging (Albany N. Y.)* **8**, 1034-1048, doi:10.18632/aging.100972 (2016).
37 Theda, C. *et al.* Quantitation of the cellular content of saliva and buccal swab samples. *Sci. Rep.* **8**, 6944, doi:10.1038/s41598-018-25311-0 (2018).
38 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408-1414, doi:10.1038/s41587-020-0591-3 (2020).
39 Holland, C. H. *et al.* Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36, doi:10.1186/s13059-020-1949-z (2020).
40 Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484-1506, doi:10.1038/s41596-020-0292-x (2020).
41 Altman, M. C. *et al.* Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children. *Nat. Immunol.* **20**, 637-651, doi:10.1038/s41590-019-0347-8 (2019).

42 Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730 e722, doi:10.1016/j.cell.2019.06.029 (2019).

Year 1 and 2 Budget

Year 1 and 2	Ordovas-Montanes	Horwitz	Yonker	Gonzalez	Glover	Majumder	Saha	de Silva	Sesay	Shalek	Sin Quee-Brown	
Personnel	$ 65,000.00	$ 65,000.00	$ 30,000.00	$ 48,000.00	$ 65,000.00	$ 65,000.00	$ 65,000.00	$ 30,000.00	$ 55,000.00	$ 65,000.00	$ 65,000.00	$ 618,000.00
Clin Supplies		$ 6,000.00	$ 6,000.00		$ 6,000.00	$ 6,000.00	$ 6,000.00	$ 12,000.00	$ 3,000.00		$ 5,000.00	$ 50,000.00
Equipment	$ -	$ -	$ -	$ -	$ -	$ 2,000.00	$ 2,000.00	$ -	$ 2,000.00	$ -	$ 2,000.00	$ 8,000.00
Travel		$ -	$ -	$ -	$ -		$ -	$ -	$ -	$ 25,000.00		$ 25,000.00
Sequencing	$ 30,000.00	$ 10,000.00	$ -	$ -	$ 10,000.00	$ 20,000.00	$ 20,000.00	$ -	$ 20,000.00	$ 10,000.00	$ -	$ 120,000.00
Subcontracts	$ -	$ -	$ -	$ -		$ -	$ -	$ -	$ -	$ -	$ -	$ -
Community Engagement				$ 28,000.00	$ 3,000.00	$ 3,000.00					$ 3,000.00	$ 37,000.00
Other Costs	$ 25,000.00	$ -	$ -						$ -			$ 25,000.00
Direct Costs	$ 120,000.00	$ 81,000.00	$ 36,000.00	$ 76,000.00	$ 84,000.00	$ 96,000.00	$ 93,000.00	$ 42,000.00	$ 80,000.00	$ 100,000.00	$ 75,000.00	$ 883,000.00
15% indirect co	$ 18,000.00	$ 12,150.00	$ 5,400.00	$ 11,400.00	$ 12,600.00	$ 14,100.00	$ 13,650.00	$ 6,300.00	$ 11,700.00	$ 15,000.00	$ 10,950.00	$ 131,250.00
TOTAL	$ 138,000.00	$ 93,150.00	$ 41,400.00	$ 87,400.00	$ 96,600.00	$ 110,100.00	$ 106,650.00	$ 48,300.00	$ 91,700.00	$ 115,000.00	$ 85,950.00	**$1,014,250.00**
Sample		BCH	Chelsea	Chelsea	Jackson	Kalkuta	Dhaka	Serrekunda			Bahamas	
Capture	Chelsea	BCH			Jackson	Kalkuta	Dhaka		Serrekunda	Frozen + Platforms	Bahamas	
Sequence	Chelsea + Jackson	BCH				Kalkuta	Dhaka		Serrekunda	Frozen + Platforms + Bahamas		
Cloud Compute	ALL					ALL			ALL	ALL		
Old budget												
Clin Supplies		6000	11000		6000	2500	6000	12000	3000		5000	51500
Sequencing	163777	163777			163777	188400	152127		152127	42000	152127	1178112
Specials	logistics director			community engagement director						training director		

Year 3 Potential Ask

Year 3	Ordovas-Montanes	Horwitz	Yonker	Gonzalez	Glover	Majumder	Saha	de Silva	Sesay	Shalek	Sin Quee-Brown	
Personnel	$ 30,000.00	$ 30,000.00	$ 15,000.00	$ 20,000.00	$ 30,000.00	$ 30,000.00	$ 30,000.00	$ 15,000.00	$ 25,000.00	$ 25,000.00	$ 30,000.00	$ 280,000.00
Clin Supplies												$ -
Equipment	$ -	$ -	$ -	$ -	$ -			$ -		$ -		$ -
Travel		$ -	$ -	$ -	$ -		$ -	$ -	$ -			$ -
Sequencing	$ 60,000.00	$ 20,000.00	$ -	$ -	$ 20,000.00	$ 40,000.00	$ 40,000.00	$ -	$ 40,000.00	$ 20,000.00	$ -	$ 240,000.00
Subcontracts	$ -	$ -	$ -	$ -		$ -	$ -	$ -	$ -	$ -	$ -	$ -
Community Engagement												$ -
Other Costs		$ -	$ -						$ -			$ -
Direct Costs	$ 90,000.00	$ 50,000.00	$ 15,000.00	$ 20,000.00	$ 50,000.00	$ 70,000.00	$ 70,000.00	$ 15,000.00	$ 65,000.00	$ 45,000.00	$ 30,000.00	$ 520,000.00
15% indirect co	$ 13,500.00	$ 7,500.00	$ 2,250.00	$ 3,000.00	$ 7,500.00	$ 10,500.00	$ 10,500.00	$ 2,250.00	$ 9,750.00	$ 6,750.00	$ 4,500.00	$ 78,000.00
TOTAL	$ 103,500.00	$ 57,500.00	$ 17,250.00	$ 23,000.00	$ 57,500.00	$ 80,500.00	$ 80,500.00	$ 17,250.00	$ 74,750.00	$ 51,750.00	$ 34,500.00	**$ 598,000.00**
Sample		BCH	Chelsea	Chelsea	Jackson	Kalkuta	Dhaka	Serrekunda			Bahamas	
Capture	Chelsea	BCH			Jackson	Kalkuta	Dhaka		Serrekunda	Frozen + Platforms	Bahamas	
Sequence	Chelsea + Jackson	BCH				Kalkuta	Dhaka		Serrekunda	Frozen + Platforms + Bahamas		
Cloud Compute	ALL					ALL			ALL	ALL		
Old budget												
Clin Supplies		6000	11000		6000	2500	6000	12000	3000		5000	51500
Sequencing	163777	163777			163777	188400	152127		152127	42000	152127	1178112
Specials	logistics director			community engagement director						training director		