bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A Multimodal and Integrated Approach to Interrogate Human Kidney Biopsies with Rigor and Reproducibility: The Kidney Precision Medicine Project

Tarek M. El-Achkar[1]*, Michael T. Eadon[1], Rajasree Menon[2], Blue B. Lake[3], Tara K. Sigdel[4], Theodore Alexandrov[5], Samir Parikh[6], Guanshi Zhang[7], Dejan Dobi[4], Kenneth W. Dunn[1], Edgar A. Otto[2], Christopher R. Anderton[7,8], Jonas M. Carson[9], Jinghui Luo[2], Chris Park[9], Habib Hamidi[2], Jian Zhou[12], Paul Hoover[10], Andrew Schroeder[4], Marianinha Joanes[4], Evren Azeloglu[11], Rachel Sealfon[12], Seth Winfree[1], Becky Steck[2], Yongqun He[2], Vivette D'Agati[13], Ravi Iyengar[11], Olga G Troyanskaya[12], Laura Barisoni[14], Joseph Gaut[15], Kun Zhang[3], Zoltan Laszik[4], Brad Rovin[6], Pierre C. Dagher[1], Kumar Sharma[7], Minnie Sarwal[4], Jeffrey B. Hodgin[2], Charles E. Alpers[9], Matthias Kretzler[2], Sanjay Jain[15]*. For the Kidney Precision Medicine Project.

Affiliations:

1- Indiana University School of Medicine, Indianapolis, Indiana
2- University of Michigan School of Medicine, Ann Arbor, Michigan
3- University of California San Diego, Jacobs School of Engineering, San Diego, California
4- University of California San Francisco School of Medicine, San Francisco, California
5- European Molecular Biology Laboratory, Heidelberg, Germany
6- Ohio State University College of Medicine, Columbus, Ohio
7- UT-Health San Antonio School of Medicine, San Antonio, Texas
8- Pacific Northwest National Laboratory, Richland, Washington
9- University of Washington, Schools of Medicine and Public Health, Seattle, Washington
10- Harvard University School of Medicine, Boston Massachusetts
11- Mount Sinai Ichan School of Medicine, New York City, New York
12- Princeton University, Princeton, New Jersey and Flatiron Institute, New York City, New York
13- Columbia University School of Medicine, New York City, New York
14- Duke University School of Medicine, Durham, North Carolina
15- Washington University in Saint Louis School of Medicine, St. Louis, Missouri

***Corresponding authors**

Tarek M. El-Achkar, Indiana University, Indianapolis, telachka@iu.edu
Sanjay Jain, Washington University School of Medicine, St. Louis, sanjayjain@wustl.edu

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author Contributions

MTE, RM, BBL, TS, TA, AS, SP, CRA, DD, EAO, SW, GZ, MJ and KD performed the ground work for the quality control group, wrote the KPMP TIS manual of procedures and generated figures. HH, JZ, RS, RM, TS, EAO, MTE, TME, PH, SP, MK, ZL and SJ generated the initial working reference marker list. CEA, TME, JBH, JL, MK and SJ led the Pilot 1 protocol. TME, VD, LB, JG, CEA, ZL, SJ and JBH developed the pathology QC tissue qualification and tissue processing criteria. JBH organized and executed the Pilot tissue collection and distribution. JC designed the SpecTrack system. CP prepared and organized the TIS manual of procedures and performed data organization services. YH led ontology development for QC metadata and knowledge standardization. BS and EA organized data integration efforts and data authentication in the data hub. KS and MS led the OMICS discussion group. SJ led the quality control group. TME and CEA led the tissue processing group. MTE and SJ led the Molecular and Pathology Integration group. MTE and SJ conceived and led the TISAC process. RI, OGT KZ, ZL, PH, BR, PCD, KS, MS, JBH, CEA, LB, JG, TME, MK and SJ conceived the integrated TIS pipeline and QC vision. TME and SJ wrote the initial draft of the paper. All authors contributed to the writing and editing of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Comprehensive and spatially mapped molecular atlases of organs at a cellular level are a critical resource to gain insights into pathogenic mechanisms and therapies tailored to the disease of each patient. Obtaining rigorous and reproducible results from disparate methods and at different sites to interrogate biomolecules at a single cell level or in 3-dimensional (3D) space is a significant challenge that can be a futile exercise if not well controlled. The Kidney Precision Medicine Project (KPMP) is an endeavor to generate 3D molecular atlases of healthy and diseased kidney biopsies using multiple state-of-the-art OMICS and imaging technologies across several institutions. We describe a pipeline for generating a reliable and authentic single cell/region 3D molecular atlas of human adult kidney with emphasis on quality assurance, quality control, validation and harmonization across different OMICS and imaging methods. Our "follow the tissue" approach encompasses sample procurement to data generation, analysis, data sharing, while standardizing, and harmonizing procedures at sample collection, processing, storage and shipping. We provide key features of preanalytical parameters, bioassays, post analyses, reference standards and data depositions. A pilot experiment from a common source tissue processed and analyzed at different institutions was executed to identify potential sources of variation, feasibility of multimodal analyses and unique and redundant features of the macromolecules being characterized by each technology. An important outcome was identification of limitations and strengths of the different technologies, and how composite information can be leveraged for clinical application. A peer review system was established to critically review quality control measures and the reproducibility of data generated by each technology before granting approval to work on clinical biopsy specimens. This unique pipeline establishes a process that economizes the use of valuable biopsy tissue for multi-OMICS and imaging analysis with stringent quality control to ensure rigor and reproducibility of results and which can serve as a model for similar personalized medicine projects.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Recent advances in biotechnology allow capturing the state of a tissue in health and disease at an unprecedented structural and molecular resolution (1). Application of these technologies at the level of the genome, transcriptome, proteome and metabolome have enabled identification of regulatory cascades and their mapping into tissue compartments at a single cell resolution (2-8). There is a unique opportunity in applying these technologies to clinical samples to decipher the complexity of tissue architecture when evaluating intrinsic organ function and its dysregulation in disease.

Kidney diseases are well-poised to be evaluated by a tissue driven approach as nephrologists routinely use kidney biopsies in a clinical setting for diagnosis and management. Using bulk and tissue compartment microdissection, first insights have been obtained by integrating histology, morphometry, transcriptional profiles and clinical outcomes primarily in glomerular diseases (9). While these studies identified therapeutic targets and disease endophenotypes, they were limited to the evaluation of glomerular diseases in routine biopsies and the analysis of tissue compartments consisting of a multitude of interacting cell types (9). Recognizing this opportunity, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) initiated the Kidney Precision Medicine Project (KPMP) with the explicit goal to bring the biomedical advances of tissue level disease process mapping to the two most devastating renal diseases, chronic kidney disease (CKD) and acute kidney injury (AKI), with a prevalence as high as 14% (37 million people) for CKD (10). In a sequential process, KPMP is establishing a framework to ethically and safely obtain research kidney biopsies from study participants with CKD and AKI, using a sophisticated tissue acquisition protocol specifically tailored to enable downstream emerging molecular genome scale analyses at tissue and cell level resolution. As KPMP sets out to establish a lasting resource both for tissue and tissue derived omics data sets, rigorous pre-

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analytical and analytical protocols with highly standardized and controlled workflows are considered an essential prerequisite for the success of the project. Here we describe a quality-controlled tissue interrogation pipeline established in the KPMP for multimodal analysis of kidney biopsies using various omics and imaging technologies. Our pipeline realizes the power of combined analyses of well vetted and curated data from different technologies to ensure rigor, reproducibility and complementarity to generate molecular atlases of healthy and disease kidneys that can impact patient care, while extracting maximum data from a limited amount of tissue. The framework established serves as a paradigm for similar atlas efforts and precision medicine projects of other organs systems and diseases (1, 3, 11, 12).

Approach

Overview of KPMP technologies: creating a molecular and functional kidney map

The KPMP encompasses a diverse set of technologies to generate a robust molecular, spatial and structural atlas (**Fig.1, 2**). A multimodal interrogation of various biomolecules including RNA and protein purposely enables a comprehensive coverage and identifies unique as well as non-captured ("blind spots") features of each technology. Redundancy among the different technologies further provides orthogonal validation that lends confidence to the discovered genes/proteins/metabolites/cell types and cell states. The overall vision enables the interrogation of the same tissue from the same study participant using many different technologies by harmonizing processing and preservation steps, to enhance data quality and to identify weaknesses and strengths of each applied technology (**Fig. 2** and see later). Following is a brief overview of technologies in the KPMP to generate a biomolecular/cell atlas of the human healthy and diseased kidney.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Transcriptomics. Several technologies are generating gene expression data at bulk, regional and single cell level for comprehensive coverage of the transcriptome for the reference (relatively healthy or pathologically normal) and disease atlases (AKI and CKD subtypes). Single nucleus (sn) (UCSD/WU) and single cell (sc) (UCSF and Premiere -Michigan/Broad/Princeton sites) RNA-seq technologies are being used for cataloging molecularly-defined cell populations for the generation a single cell kidney atlas (13-15). Each of these technologies uses different approaches in sample preparation, dissociation or processing that have unique advantages and disadvantages (see TIS protocol on https://kpmp.org/researcher-resources/). Given the potential to introduce artifacts or miss cell types due to processing of these specimens, concurrent bulk RNA expression analysis on undissociated tissue or regional laser microdissection (LMD) (IU/OSU) is performed (16-18). Using the LMD technology, transcriptomic signatures are identified for seven sub-segments of the nephron including the proximal tubule (PT), thick ascending loop of Henle (TAL), distal convoluted tubule (DCT), collecting duct (CD), as well as compartment-specific signatures for glomeruli, the tubulo-interstitium (TI), the interstitium (without glomeruli or tubules), as well as a bulk cross-section of the entire biopsy. LMD transcriptomic signatures can serve as an important independent validation measure that provides regional spatial context to cell populations discovered using single cell technologies. To complement the mRNA signatures obtained from the single event technologies and regional transcriptomics, miRNA sequencing of bulk cross-sections from the same OCT-embedded core are sequenced specifically for small RNA (IU/OSU) (19).

Proteomics. Two different approaches were planned in the KPMP to generate reference and diseased kidney proteome. The regional/segmental approach by the IU/OSU group combines the collection of specific regions of the kidney using LMD with quantitative proteome analysis using state-of-the-art HPLC/mass spectrometry (MS) instrumentation to generate agnostic global

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

profiles of the glomerular and tubulo-interstitial compartments with high sensitivity and reproducibility (20). The UCSF group is employing recently developed nanoscale proteomics analysis in which a few cells can be processed to generate cell-specific protein profiles (21). Combining this technology with regions isolated by LMD provides regional and spatial definitions of generated proteome data. Both technologies additionally provide bulk proteomics data on tissue sections allowing cross platform/site analysis.

Metabolomics. The UTHSA/PNNL group will generate spatial metabolomics measurements by using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI), an approach that has been utilized previously for kidney molecular imaging (22, 23). Here, fresh-frozen tissues are cryosectioned, mounted on optically transparent and electrically conductive slides, coated with an organic matrix that assists in facilitating desorption and ionization of endogenous molecules, and serially probed with a laser to attain mass spectral information at predefined locations. They employ two different platforms to generate metabolite profiles that are designed for cross validation and complementarity. An important aspect is that data are processed using the METASPACE platform (24, 25) (a tool developed by the EMBL portion of this group), which is an automated molecular annotation engine that enables data visualization and co-registration with other optical images.

2D and 3D Imaging. There are several imaging technologies within KPMP that inform on precise 2D and 3D expression relationships of biomolecules, cells and structures with varying degree of multiplexing, spatial resolution and tissue preservation conditions **(Fig. 2).** The UCSD/WU group will develop DART-FISH to delineate single cell mRNA expression of several hundred transcripts in each cell at high resolution in 3D using high resolution confocal microscopy. The UCSF and PREMIERE sites will use miFISH to generate simultaneous gene

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and protein expression data at the single cell level in 2D space (26). The IU/OSU site will utilize large scale 3D confocal imaging of longitudinal kidney thick sections probing concurrently for 8 different targets using antibodies or fluorescent small molecules, followed by tissue cytometry analysis (27, 28). Prior to any staining, this technology uses label-free imaging to determine tissue integrity using endogenous fluorescence and quantify collagen deposition using second harmonic generation. The resulting image volumes undergo cytometry analysis with a customized Volumetric Tissue Cytometry and Analysis (VTEA) software. This approach allows the interactive exploration of the image volumes, as well as quantitative analysis of the abundance, distribution, and other 3D spatial features of interest for various cell types, based on supervised and unsupervised analytical approaches (29). The UCSF site will use Co-detection by indexing (CODEX) (30), an orthogonal approach to perform 2D profiling *in situ* on up to 30 antigens at single-cell resolution on a single tissue section using antibodies labeled with unique oligonucleotide tags ("barcodes"). The labeling and detection is done in an iterative manner in groups of two or three targets per cycle by fluorescent dyes labeled with oligonucleotide sequences ("reporters") corresponding to a given subset of antibody barcodes; in each cycle the previous reporters are removed and a new set is introduced. The montage image consisting of all signals from each cycle is analyzed for marker intensity distribution across the tissue section, for cell count of a given cell population defined by marker sets and for spatial relationship between cell populations of interest in the specimen (30).

Guiding principles for multimodal quality-controlled tissue interrogation

Biomolecular and imaging data to build a kidney atlas in KPMP consists of partnerships between multiple institutions as Tissue Interrogation Sites (TISs) to generate various single cell/region OMICS (transcriptomics, proteomics and metabolomics) and 2D and 3D high resolution kidney imaging data using technologies described above (**Fig. 1, 2**). These multi-scalar imaging and

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

single cell/region omics technologies provide an unprecedented resolution, big data for interpretation and insights into biological processes. With these attributes comes the challenge of compounding errors as the data are dependent on multiple processes and steps that begin from specimen procurement to data generation and analysis. When data are generated from different institutions and on archived tissues there are several other sources of random and technical variations that confound the outcomes and impact biological reproducibility and interpretation of data. An additional challenge is maximizing the application of these sensitive, big data technologies to clinical biopsies with limited tissue available for research. Motivated to overcome these challenges in KPMP, our approach described here aims to standardize and harmonize, where possible, the entire process of tissue procurement all the way up to analysis. Key factors considered were economizing tissue usage, maximizing preservation and use for multiple technologies, and documenting quality of intermediate steps with clearly defined quality assurance and control criteria. Central to this theme was the application of rigorous procedures for transparency, technical and biological reproducibility across multiple sources of tissue procurement, analysis and interpretation. Below we provide an overview of the approach, examples of different technologies including current quality control metrics, procedures for harmonization and standards to be met for interrogating human kidney biopsies.

Overview of the strategy: "Follow the tissue"

An overall goal of our approach is to derive synergistic and integrated information on the biological state of tissue in health and disease. Anticipating that the tissue would come from different sites, our approach was to develop a tissue processing pipeline that can be easily implemented at the bedside and applied to multiple state-of-the-art interrogation technologies. We required that each technology independently investigates and analyzes adult human kidney tissue in a reproducible and rigorous manner. The goal for this approach was to demonstrate

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

feasibility, perform validations, harmonize procedures and enable integrated analysis with other technologies executed in KPMP (**Fig. 2**). This process posed several challenges at multiple steps in the "follow the tissue" process including variable standards in collecting metadata, procurement, processing, sample preparation, analytical parameters, analysis and data sharing and deposition. To tackle these challenges, we formed working groups for tissue processing that included expertise in OMICS, imaging, and pathology. These groups addressed quality control measures across four main categories: 1) Participants, 2) Tissue, 3) Assays and analysis, 4) Data hub (**Fig. 3**). Each category encompassed various components with parameters for quality assessment and control. The working groups met weekly through web-conferences and offline with concrete action items and delegated tasks to identify critical parameters and data to be collected in each of the 4 categories. These meetings were open to all KPMP members and meeting minutes were documented at the KPMP management website in "Basecamp" (www.basecamp.com). This process enabled easy access to archived documents and meetings, and active discussions among members. A critical aspect was formulating and capturing well-vetted relevant metadata in each of these categories to enable the interpretation of molecular discoveries among the underlying biological variations related to healthy and disease phenotypes (**Table 1)**. We established benchmarks for the entire process, and a clear approach for data visualization and sharing for various types of users. Once the overall vision or "blueprint" was established, the pipeline was pressure-tested with a pilot project using adult human kidney tissue (see later). We describe below the individual steps of this workflow.

Subject Metadata. Detailed relevant participant associated metadata including demographics, clinical and pathological diagnoses, comorbidities, medications, social histories and laboratory data were identified through a team effort of all the recruitment sites and TISs and will be described elsewhere. This was important to interpret variations in data due to contributions of

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

patient attributes. For example, sex differences, age, race or medications such as diuretics or antihypertensives can contribute to changes in molecular or cellular distribution of transporters in the kidney. An example of the main standardized patient metadata fields used by all the sites is provided (**Table 1**). These metadata fields are being modeled and represented using the Kidney Tissue Atlas Ontology (KTAO)(31), and Ontology of Precision Medicine and Investigation (OPMI) (32), two open source biomedical ontologies that are being developed by the collaborative KPMP and ontology communities.

Tissue Preanalytical Considerations. The output of OMICS and imaging technology results is intimately related to tissue handling and preservation attributes, which if not controlled can contribute to technical variations and decreased data quality. Since there are more than 10 tissue interrogation technologies but only sparse tissue availability (biopsy), the methods in which tissue is collected, preserved, processed, stored, shipped, analyzed and recorded were harmonized, whenever possible, to minimize technical variations (some discussed below). Several preanalytical parameters for quality control related to specimen handling, processing, preservation, orientation, quantity used, shipping and storage were identified to minimize these concerns (**Table 1**). An infrastructure to track specimen movement from origin to specimen processing sites was developed (SpecTrack). A key feature was the ability to record deidentified specimens and documentation of the time and state in which the specimens are shipped and received accompanied with a photograph of the contents. All the sites were required to show successful use of this system prior to being qualified to receive specimens. An important consideration in designing the pipeline was to assess the quality and composition of the tissue being analyzed to best interpret the molecular outcomes. While some technologies depended on complete dissociation of tissue (scRNA-seq) others had the opportunity to register tissue composition before dissociation or analysis. As such, guidelines were developed by the KPMP

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pathologists and TIS investigators for high level quality assessment of composition and integrity of tissue sections and included relative proportions of cortex, medulla, glomerulosclerosis and features compromising integrity such as necrosis and hemorrhage (**Table 1**).

Tissue Analytical and Assay Considerations. There were three main tenets we focused on: 1) Assay metadata, 2) Assay quality assurance (QA) parameters, and 3) Assay quality control (QC) parameters. We ensured careful attention was made from the beginning to define and record all the metadata associated with the assays, with the goal of transparency and reproducibility. QA is linked to understanding and applying the best practices recommended in the field for that particular technology, using a specific instrument set(s), protocols or platform. The reliance on data from the manufacturer or post-marketing analysis when available is essential. We ensured that all platforms used in data production are optimized to produce the best possible results or operating under *bone fide* core facilities. For the QC component, we expected to meet and exceed a set of criteria guaranteeing that the assay works properly; these criteria also set a benchmark to give reproducible data for building the kidney atlas.

Considering the complexity of assays and data outputs across technologies, our goal was to harmonize metadata collections, assays, instrumentations and *post hoc* analyses for similar biomolecules where possible. During sample preparation for assays and analysis, quality control parameters and minimum attributes that are relevant for the performance of the assay were identified and common terms were used for similar types of technologies (RNA or protein or imaging). Key parameters needed to be identified for each technology to be captured during assays that could be tracked to monitor quality at different time points throughout the respective pipeline. Each technology was expected to come up with concrete criteria for quality control measures and to demonstrate pass-fail rates and reproducibility in pilot experiments (see later).

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Furthermore, within each technology, implementation of measures that allow detection and control of batch effects and assay drift were also incorporated.

We followed the concept of building an iterative marker list derived from published data and data generated from the KPMP to qualify the identity and composition of the tissue being interrogated, validate and optimize tissue processing pipelines and identify regions or cell types for integrative quality check and analysis to build the kidney atlas. Our initial list (made in 2018) of a subset of cell type markers mainly relied on rodent studies and bulk RNAseq data with corresponding evidence from the human protein atlas (**Supplemental File 1)** (33-42). Later iterations of the potential cell types/states were heavily dependent on the data generated from the human kidney specimens derived in part from the pilot project (see below). Similarly, for imaging studies a number of parameters were established to best standardize the formats of image acquisition, analysis and deposition (**Supplemental Table 1**).

Data quality check, visualization and sharing. We placed additional quality checks for the generated data by establishing a non TIS directed team in “data hub” where the investigators could seamlessly transfer data upon passing their QC checks. Among the functions of the “data hub” are: 1) the examination of the associated metadata for completeness, 2) independent analysis of the data for passing quality control thresholds, 3) enhancing data availability to other KPMP sites for integrated analysis and quality check and 4) planning for public sharing. Indeed, an essential component of data output is making it accessible to the public. In this regard, the KPMP data hub is tasked with a team dedicated for building tools for summary analysis and visualization of the integrated results generated by the various technologies.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Anticipated challenges and how to handle them: harmonization vs complementarity, and the need for a pilot experiment

Based on the overall vision of the "follow the tissue" pipeline, many challenges were anticipated because of the extensive breadth of the technologies applied to interrogate common specimens. How do we reconcile outputs from these various approaches? Is it possible to store and ship tissue to distant sites without compromising it for interrogation by different technologies? How do we benchmark the quality control parameters for evolving technologies, while at the same time economizing and maximizing the use of precious tissue? Within each technology, can we define steps in the analytical pipeline that could be harmonized with other technologies without undermining the uniqueness and value presented by each technology alone? Can we strike the appropriate balance between harmonization and complementarity? Indeed, such balance would solidify the overall approach and potentially become a strength and a staple of KPMP to promote discovery. To tackle these challenges and demonstrate the feasibility of the analytical pipeline, we recognized that the "follow the tissue "approach required testing in a consortium wide experiment. Therefore, a KPMP pilot experiment was designed and executed as described below.

Testing the quality control pipeline: a consortium wide pilot experiment:

Rationale: The objective of this experiment was to use a same source kidney specimen to pressure-test our "follow the tissue" approach on a consortium level. We aimed to assess our ability to 1) standardize tissue processing/handling, storage and shipping steps; 2) establish feasibility and validate the Q/A-Q/C parameters for all the technologies in the interrogation pipeline; 3) compare, when applicable across sites, the performance of molecular interrogation and identify sources of variabilities and concordance; 4) lay out a blueprint for harmonization and complementarity across technologies; 5) identify gaps and weaknesses in the interrogation pipeline. An important outcome was to define a protocol that is harmonized across technologies,

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and that could ultimately be used for interrogating biopsies from patients. This protocol is also expected to economize the use of limited tissue material for diagnostics and research.

Design:

The approach for the pilot was to collect tumor-free kidney cortex from nephrectomy specimens from the University of Michigan tissue collection center, preserve the tissue in different types of media according to the needs of the various TISs and distribute to each TIS for testing feasibility, validation and identifying the quality control metrics for their respective technologies. Contiguous serial sections (approximately 1cm x 2mm x 2mm) in the shape of rectangular cuboids were cut for processing and preservation (FFPE, fresh frozen OCT, flash frozen in liquid nitrogen, Cryostor, RNAlater) and shipped to each TIS designated by a code (**Figure 4 and Supplemental Figure 1**). In total, 6 different nephrectomy specimens were processed as described above and used by all the TISs. Hence, not only each site had access to the same tissue source, but there were also 6 biological replicates distributed for the purpose of testing reproducibility, as discussed below.

Quality control outcomes and observations based on the pilot experiment:

1) Tissue procurement/preservation - Preanalytical parameters: The following outcomes were directly derived from this pilot experience:

(A) *Better definition of the metadata associated with tissue procurement, preservation, integrity and composition*. We identified commonalities between tissue procurement, processing, assessment and storage that enabled the use of similar conditions for multiple technologies (**Table 1).** For example, snRNAseq, 3D-mass cytometry, LMD transcriptomics, LMD proteomics, mDroscRNAseq, spatial metabolomics, miFISH and DART-FISH could all use fresh frozen OCT blocks (**Tables 2, 3** and detailed in TIS manual of operations at www.kpmp.org/resources).

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(B) *Ontology-based metadata modeling and representation.* Such an approach will likely facilitate the definition of the metadata and make the links between different metadata types more meaningful and machine-interpretable, supporting advanced data analytics and knowledge discovery.

(C) *Real time testing of specimen tracking using the SpecTrack software.* This live tracking of the tissue revealed weaknesses in the pipeline and allowed improvements, including better documentation of tissue and temperature states of shipments and appropriate packaging materials (**Supplemental Figure 2**).

(D *Effect of shipping and best practices establishment*. To determine the effect of shipping on tissue quality, an assessment of RNA integrity was performed on bulk tissue preserved in RNAlater using independent RNA preparation methods at two different sites. All the bulk RNA samples (total 12, 6 nephrectomy samples in RNAlater shipped to each site) were sequenced at a central site. These results showed strong correlation among adjacent tissue samples from the same subject for all 6 subjects and established the shipping conditions that do not adversely affect tissue state as measured by RNA expression and integrity analysis (**Figure 5***).*

(E) *Initial processing at the TISs.* This experiment also provided an opportunity to examine the initial processing steps at each TIS, to explore the potential of standardizing common procedures, whenever possible. This resulted in the implementation of common procedures at each site, which were incorporated in the KPMP TIS manual of procedures (www.KPMP.org). For example, this pilot experiment identified the need to obtain histology sections flanking areas of interrogation within the tissue, to inform on the state, composition, and orientation of the tissue. This process also allowed the same OCT block to be exchanged by two interrogation sites to perform successful molecular interrogation simultaneously with 3 different techniques (**Figure 5**). The pilot experiment was also crucial to verify, validate and expand the metadata

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

variables that needed to be captured for faithful documentation of the tissue journey from harvesting to interrogation.

2) Analytical Q/C parameters for each technology based on pilot 1: One of the main goals of this experiment was to test and validate the quality control parameters for each technology in each site. This experiment also presented a venue to test for both biological and technical reproducibility. There were in total 6 different nephrectomy specimens distributed to each site, with a substantial amount of tissue from each specimen. As such the TISs were able to optimize their technologies, check feasibility and in parallel apply the refined methods to their locally collected kidney samples. Repeat testing on a single specimen provides a framework for technical reproducibility; the use of tissue from different donors and tissue from different sources (pilot and local samples) ensured testing the methodologies for rigor and biological reproducibility. The Q/C parameters adopted by each technology based on this pilot experiment are summarized in **Tables 2, 3** and **4**.

Solidification of Q/C parameters was also performed by cross-validation with existing data/standards or by cross-validating outcomes from various technologies performed on the single source kidney tissue provided in this pilot. This cross-validation could be in the form of concordant readouts, such as detection of the same molecules/metabolites in the same samples using different technologies. In addition, orthogonal validation could also occur using concordant "derived" readouts such as pathway analyses. For example, the TIS technologies can detect different genes/molecules/metabolites, but these molecular entities can be part of the same signaling pathway. Examples of orthogonal validation approaches are shown in **Figure 5** and will be presented in detail in a separate manuscript.

3) Post-analytical outcomes: In addition to the cross-validation benefits discussed above, examining the outputs from various technologies promoted integration efforts and helped

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

determine the extent of complementary information provided by each technology. This ensures that a comprehensive cellular and molecular converge is provided by the consortium to make robust kidney atlas and provide a platform for discovery.

This post-analytical exercise also provided the opportunity for further metadata harmonization, at the various levels of tissue processing, analytics and analysis. Additionally, parameters for diagnostic features, composition, and integrity of the tissue that are applicable to all the TISs were further refined and led to a protocol for interrogating patient biopsies in the KPMP described in a comprehensive Pathology protocol document (https://kpmp.org/researcher-resources/).

An additional important outcome was that significant amount of gene and protein expression data were generated from the pilot samples. These data collected from multiple sites provide an initial view of cellular diversity in the human kidney (**Figure 6, Table 5)**. The analysis also revealed stress states related to processing of tissue and underlying pathology that could not have been predicted from gross evaluations in presumably healthy tissue. In fact, some novel discoveries have already emerged in the initial version of the kidney atlas from the Pilot project (13).

Identification of gaps and improvement of the process

An area of priority identified during the integration efforts of the large OMICS and imaging datasets was the need to establish benchmarks for the nomenclature of cell types, regions and associated genes, proteins and metabolites for reference and disease atlas and various injury states. A promising methodology of analysis that could link multiple technologies is a cell-centric approach, whereby the outputs can reflect changes at the cell level in a tissue specimen. This was essential as several groups are investigating the single cell transcriptome or proteome of the kidney but there is lack of conformity regarding nomenclature and annotations. However, this analytical process requires an initial definition of cell types based on a set of criteria, such as: gene expression (RNA and protein), cell state (baseline, stress, injury), spatial localization and

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

associations, among others. The pilot studies generated an initial working list delineating the complexity of cell types and a subset of associated marker genes in the adult human kidney which could serve as a starting point for kidney OMICS and imaging studies for classification of cell types and states and harmonization with recent renal tubule epithelial cell nomenclature (**Table 5**) (42). An ontological representation of the cell markers has been initiated to seamlessly link gene, cell type and spatial tissue location at an integrative semantic level (31).

After establishing qualification criteria for tissue processing and the analytical process, a key question was if the technologies are ready to interrogate the limited clinical biopsy specimens. Therefore, it was unanimously perceived that a mechanism needs to be established to firmly vet and anchor each technology for quality control, rigor and reproducibility before handling any prospective precious biopsy tissue. This resulted in the establishment of a peer review technology approval process, which will be discussed below.

Implementation of best practices to perform tissue interrogation on KPMP biopsies: the TISAC process:

Approval of TISs to receive biopsies for interrogation. To rigorously evaluate each technology and eliminate self-approved bias by each TIS we established a Tissue Interrogation Site Approval Committee (TISAC). This committee evaluates all technologies prior to approval to perform studies on KPMP biopsy samples. The committee is composed of representatives from the central hub, DVC, recruitment sites, other TISs, and external *ad hoc* members as required to provide sufficient expertise to review the technology. The committee organizes webinars for the TISs to receive feedback regarding technology readiness. A portfolio is submitted by a TIS for each technology which addresses the protocol, quality control metrics, sample handling, batch effects, assay drift and complementarity. These elements are summarized in **Figure 7** and **Supplemental file 2 [TIASC checklist]**. Three primary reviewers formulate critiques after

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reviewing the portfolio, consult among each other and then the entire portfolio is discussed in a TISAC review call with representation from the TISs (except the TIS under review), non-TIS members and the NIDDK. The TISAC provides constructive feedback to enhance the reproducibility and complementarity of each technology, as well as identify areas that require additional supporting data. Once satisfied with a given technology's readiness, the TISAC recommends approval and notifies the Steering Committee. The TISAC provides their report to the NIDDK and the KPMP external expert panel who ultimately approve the technology and TIS for receipt of patient samples. Each TIS is further expected to report on the state of their technology at least annually and sooner if there are any modifications in the protocol used.

Ongoing progress, challenges and future outlook

Designing and implementing a quality control pipeline for the various technologies in KPMP required an *a priori* conceptualization of the scope, type and usefulness of the data that led to the "follow the tissue" pipeline concept with quality control components embedded throughout presented here. In addition to the expertise of investigators within and outside KPMP, we were also guided by the interaction with large national and international consortia. The validated pipeline that began as a concept provides a paradigm for similar efforts in healthy and disease atlas projects. The vision of KPMP is to evolve dynamically by incorporating ongoing progress in science, technology and patient care. Therefore, an important component of KPMP is an ability to identify opportunities to improve the quality control process, tackle challenges that are presented when changes are introduced by evolving or newer technologies, and mitigate potential threats or unforeseen errors. Examples of ongoing areas of development or topics that could present a challenge in the immediate future are described below:

- Ontology representation of metadata types, cell markers, assays, and assay components to support harmonization. The urgency of this task has been recognized and has been the focus of dedicated workshops. In addition to linking the data, ontologies will be essential to

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

link all the quality control elements within the data and metadata. The KPMP QC and ontology working groups are collaborating and making progress in "ontologizing" and linking different data types, cell markers, assays and assay components. KPMP is expected to define new landscape related to kidney disease specific ontologies that will be the subject of a separate publication.

- DVC integration and dissemination of results. This area is currently in development. It is expected that data (raw and processed), metadata and all Q/C elements will be publicly available for all types of users in a way that is easy to query, access and interpret.
- Incorporation of external data. The ability of the interrogation pipeline with its various technologies and Q/C elements to interact with external data will be important to extend the relevance and reach of KPMP discoveries.
- Future proofing and addressing technology drifts.

A few of the key major future challenges are outlined in **Table 6**.

Conclusions

With the implementation of a multimodal and integrated pipeline for molecular interrogation of kidney biopsy specimens, the KPMP experience will be unique in its goals to set high standards for quality control, rigor and reproducibility. Vetted technologies participating in KPMP will undergo careful scrutiny to comply with these goals of quality control, while at the same time allowing a dynamic and iterative approach that promote improvement and transparency. In doing so, KPMP could become a model for other national and international efforts that also seek to decipher human disease and build a dynamic tissue atlas. KPMP's ultimate goal is to improve patient care and cure kidney disease.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

The Kidney Precision Medicine Project is supported by the National Institute of Diabetes and Digestive and Kidney Diseases through the following grants: UH3 DK114923, UH3 DK114920, UH3 DK114933, UH3 DK114937, UH3 DK114907, U2C DK114886. We thank the KPMP patient participants, Recruitment sites, Central Hub and all the TISs for many valuable discussions and feedback towards the QC efforts. We are grateful to the KPMP Publications and Presentation committee for suggestions and review of this manuscript. A complete list of all KPMP members can be found at kpmp.org.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Preanalytical parameters for rigor, reproducibility, quality assessment and control.

Subject	Specimen Procurement	Specimen Morphology	Specimen Storage/Shipping
Age	Procedure data	Gross assessment	Storage time
Race	Procedure Type (U/S, IR, Surgical)	Integrity (intact, fragmented)	Storage temperature
Sex	Location/Site	% Cortex	Storage medium
Height	Procurement time	% Medulla	Storage container
Weight	Specimen collection media	Microscopic assessment	Shipping time
BMI	Tissue transport time to lab	# Glomeruli	Shipping container
Clinical labs (blood, urine)	Transport temperature	% Global glomerulosclerosis (age corrected)	Shipping image
eGFR	Preservation media/condition	Necrosis	Receiving time
Clinical diagnosis	Supplies/Reagent details	Hemorrhage	Receiving image
Pathology diagnosis	Processing time	Autolysis	Storage time receiving
Comorbidities	Size	Tubular atrophy	Storage temp receiving
Medications	Image	Interstitial fibrosis	
Social history		Interstitial inflammation	
Family History		Image (gross, HE, PAS)	
		Freezing artifacts	

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Standardization and Quality Control Parameters for RNA Omics Technologies. "*" denotes technologies approved by TISAC for interrogating patient biopsies (Oct 2019).

Category/Site	**UCSD-WU**	**PREMIERE**	**UCSF**	**IU-OSU**	**Standardized QC**
Technology	**snRNA-seq***	**scRNA-seq***	**mDroscRNA-seq**	**LMD mRNA* subsegmental and bulk miRNA**	**All RNA OMIC**
Sample Preservation/Storage/Processing/ Transport Parameters	• Fresh frozen OCT • -80 °C • Cryosection (5-20 cu mm), RNAlater • Nuclear prep (>50K) • Dry ice	• Fresh frozen Cryostor • -80 °C • Dissociation (Liberase) • Dry ice	• Fresh frozen Cryostor • -80 °C • Dissociation (Liberase) • Dry ice	• Fresh frozen OCT • -80 °C • mRNA: Cryosection, dissect subsegments < 2 h, >500Ksqm • miRNA: Cryosection 20-60 µm • Dry ice	• Pre-analytical • Tissue procurement • Tissue processing (preservation, path assessment) • Shipping (dry ice) • Supplies • Storage (-80 °C)
Assay (RNA QC, library preps, Instrument, Sequencing)	• DV200 > 25% • Lib size (200-1000bp) • Bioanalyzer 2100 • Drop-seq or 10X Chromium • Illumina • Paired end (30x100)	• Lib size (200-1000bp) • Bioanalyzer 2100 • Illumina • 10X Chromium • Paired end (30x100)	• Lib size (200-1000bp) • Bioanalyzer 2100 • 10X Chromium • Illumina • Paired end (30x100)	• DV200 > 25% • Lib size (200-1000bp) • Bioanalyzer 2100 • RNA ref std (Stratagene) • Illumina • Paired end (30x100)	• RNA quality (DV200>25%; RIN for bulk > 6.0) • Library size (200-1000bp) • RNA ref std for bulk (Stratagene) • Instrument • Sequencing platforms
Analytics (QC filters, artifacts, softwares, thresholds)	• >400<5000 genes/nucleus • Seurat V3 or Pagoda 2 • UMAP cluster • Reference markers for annotations • >30 cell/integrated clusters • >50 QC nuclei/sample • GRCH38	• >500<5000 genes/cell • MT<20% or <50% (Seurat V2/V3) • UMAP cluster • Reference markers for annotations • >30 cell/integrated clusters • >100 QC cell/sample • GRCH38	• >500<5000 genes/cell • MT <50% (Seurat V3) • UMAP cluster • Reference markers for annotations • >30 cell/integrated clusters • GRCH38 • >100 QC cell/sample • GRCH38	• Read count > 10 over 50% of samples for at least 1 subsegment • >1 million reads/sample • Edge R • Reference markers for annotations • GRCH38	• >400/500<5000 genes/cell (sn/sc) • Seurat V3 • Reference genome GRCH38 • UMAP cluster • Reference markers for annotations • >30 cell/integrated clusters • >50/100 (sn/sc) datasets per sample
Data deposition	• Fastq, bam • Count matrix R Objects	• Fastq, bam • Count matrix R Objects	• Fastq, bam • Count matrix R Objects	• Fastq, bam • Count matrix R Objects, csv, txt	• Fastq, bam • Count matrix R Objects • csv, txt

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3. Standardization and Quality Control Parameters for Proteomics and Metabolomics Technologies.

Category/Site	UCSF	IU-OSU	UTHSA-PNNL-EMBL	Standardized QC
Technologies	**nscProteomics**	**LMD regional proteomics**	**Spatial Metabolomics**	**All Proteomics and Metabolomics**
Sample Preservation/Storage /Processing/ Transport Parameters	• Fresh frozen OCT • -80 °C • Cryosections • Dry ice	• Fresh frozen OCT • -80 °C • Cryosections • Dry ice	• LN2 snap frozen • -80 °C • Cryosection • Dry ice	• Pre-analytical • Tissue procurement • Tissue processing (preservation, path assessment) • Shipping (dry ice) • Supplies • Storage (-80 °C)
Assay	• adjacent sections > 10 cells per region • Orbitrap Fusion (< 3 ppm calibration accuracy) • Peptide standards	• 10,000-20,000 glomerular or tubular cells • Orbitrap Fusion (< 3 ppm calibration accuracy) • Peptide standards	• Entire tissue section, bulk analysis on remaining • Orbitrap QE and FTICR-MS (< 3 ppm calibration accuracy) • Commercial tune mix	• Tissue sectioning • Cell /region isolation • Orbitrap Fusion (proteomics) • Peptide standards (proteomics)
Analytics	• >2000 proteins per 10 cell prep • > 6aa • <4600Da peptide mass • Missed cleavage for peptides = 2 • < 1% FDR • > 97% correlation for peptides in replicates • Reference markers for annotations	• > 3000 proteins identified • < 1% FDR • Reference markers for annotations	• >100 annotated metabolites per section (<20% FDR) • Detection of standard metabolites • Detection of location specific metabolites • Histology correlation	• > 2000/3000 proteins (nsc/LMD) • < 1% FDR (nsc, LMD) • Standard metabolites and renal metabolites
Data deposition (file types)	• .raw, .csv	• .raw, .csv	• MALDI-MSI: .d/.mis, .xml, .imzML, .ibd, raw, csv • LC-MS/MS: .raw, .mzML	• raw, csv (proteomics) • raw (LC-MS/MS) • d/mis, imzML, ibd, raw, csv (MALDI-MSI)

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 4: Label-based imaging QC. "*" denotes technologies approved by TISAC for interrogating patient biopsies (Oct 2019).

	IU-OSU	UCSD-WU	UM-Broad-Princeton	UCSF	
Technology	3D Tissue Cytometry*	DART-FISH2	ISH	mIFISH	CODEX
Probe Targets Per Section	8 Immunofluor targets	>50 mRNAs	3-10 mRNA	20 mRNA & proteins	30 proteins
2D or 3D Imaging?	3D	3D	2D	2D	
Spatial Resolution	0.5-1 µm	1 µm	single cell	1µm/single cell	
Cell type and structure mapping	Validated antibodies and fluorescent small molecules known to bind specific cell types	nuclear stain, lectin labeling for PT or CD, autofluorescence	cell specific RNA and histological stain in sequential plane	Anchor (structural) and immune cell and interstitial compartment mapping at single cell resolution with direct H&E correlation	
Area/Volume of Imaging Region	Entire biopsy length (1-1.5cm) x 50µm thickness	10-20 micron sections of 1 mm^2	Entire biopsy length (1-1.5cm) x 5 µm thickness	Entire biopsy (20-25 mm^2)	
Dimensions of Image Voxels	0.5 x 1µm x 1µm	0.15 x 0.15 x 0.3 µm	N.A.	N.A.	
Amount of Tissue	Minimum: One 50 um thick section (imaging/staining)	10 x 10 µm thick sections	5 µm thick section for each probe target	One 3µm-thick FFPE section for each 21-plex mIFISH stain. Two 3µm-thick FFPE sections for a positive/negative ISH control.	One 5µm-thick OCT section for each 32-plex CODEX stain
Pre-imaging QC	-Primary Ab validation -Pre-staining with secondary antibodies as a negative control -AF/SHG (tissue quality) -16 channel pre-stain imaging	-Validation of generated probes -Maintenance of fluorescent imaging system	In development	-Instrument: Auto-calibrated with each use -Tissue: H&E to assess tissue integrity -Antibody validation: IPOX/IIF/orthogonal	-Instrument: standardized imaging of IF calibration slides at regular intervals -Tissue: H&E-stained sections to assess tissue integrity/freezing artifacts -Oligo-labeling: comparison of the staining across CODEX, indirect IF, and IPOX platforms
Acquisition QC	-Reuse acquisition settings -Point Spread Function -Signal to noise ratio	-Signal to Noise Ratio -Number of rolonies decoded -Number of genes decoded	-Negative control: DapB -Positive control: housekeeping genes	-On-slide tissue microarray controls -Built in software function to set exposure times -New antibody lots compared to "gold standard" -- epifluorescent images for batch consistency -ISH: positive & negative control probes	-Labeling-detection: on-slide tissue microarray controls -Imaging: saturation based on histogram -Cell clustering by intensity gating: visual assessment of the raw image
Analysis QC	Unmixing (labeled beads at known ratios) Segmentation quality (F1 scores)	-Misclassification rate (accuracy of barcode decoding) -Number of rolonies per cell	In development	Visual inspection of analyzed images: segmentation & analyte detection accuracy	Cell clustering by intensity gating & visual assessment of the raw image
Imaging and Analysis Throughput	2-3 samples per week	1 section per week	In development	8 sections/week (mIFISH)	2-3 sections/week (CODEX)

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 5: Cell types and associated markers from KPMP Pilot 1 transcriptomic studies. Asterisk denotes genes detected by more than one technology. *Italics*, genes detected by a single technology.

Structure/Region	Sub structure/Sub region	Cell Type	Abbreviation	Subset of Marker Genes	Pertinent negatives/comments
Renal Corpuscle	Bowman's Capsule	Parietal epithelial cell	PEC	*CRB2*, CLDN1**	
	Glomerulus	Podocyte	POD	*NPHS2*, PODXL*, NPHS1**	
		Capillary Endothelial Cell	GC-EC	*EHD3*, EMCN*, HECW2*, FLT1*, AQP1**	
		Mesangial Cell	MC	*POSTN*, PIEZO2*, ROBO1*, ITGA8**	
Tubules	Proximal Tubule	Proximal Tubule Cell (general)	PT	*CUBN*, LRP2*, SLC13A1*, ALDOB*, GATM**	
		Proximal Convoluted Tubule Cell Segment 1	PT-S1	*SLC5A2*, SLC5A12**	There is overlap among the segments
		Proximal Tubule Cell Segment 2	PT-S2	*SLC22A6**	
		Proximal Tubule Cell Segment 3	PT-S3	*PDZK1IP1*, MT1G**	
	Loop of Henle, Thin Limb	Descending Thin Limb Cell (general)	DTL	*CRYAB*, VCAM1*, AQP1*, SPP1**	CLDN10 low
		Ascending Thin Limb Cell (general)	ATL	*CRYAB*, TACSTD2*, CLDN3**	AQP1 low to none
	Loop of Henle, Thick Limb	Thick Ascending Limb Cell (general)	TAL	*SLC12A1*, UMOD**	SLC12A3 low to none
		Cortex-TAL cell	C-TAL	*SLC12A1*, UMOD**	
		Medulla-TAL cell	M-TAL	*SLC12A1*, UMOD**	
		TAL-Macula Densa cell	TAL-MD	*NOS1*, SLC12A1**	
	Distal Convolution	Distal Convoluted Tubule Cell (general)	DCT	*SLC12A3*, TRPM6**	
		DCT type 1 cell	DCT-1	*SLC12A3*, TRPM6*	SLC8A1, HSD11B2 (low to none)
		DCT type 2 cell	DCT-2	*SLC12A3*, SLC8A1*, HSD11B2*	Has CNT and DCT signature
	Connecting Tubule	Connecting Tubule Cell (general)	CNT	*SLC8A1*, CALB1, TRPV5*	SLC12A3 low to none
		CNT-Principal Cell	CNT-PC	*SLC8A1*, AQP2*, SCNN1G**	
		CNT-Intercalated Cell	CNT-IC	*SLC8A1*, CA2, ATP6VOD2**	
		CNT-IC-A	CNT-IC-A	*SLC8A1*, SLC4A1*, SLC26A7**	
		CNT-IC-B	CNT-IC-B	*SLC8A1*, SLC26A4*, SLC4A9**	
	Collecting Duct	Collecting duct (general)	CD	*GATA3**	GATA3 may be in subpopulation of DCT, CNT and vSMC/P. SLC8A1, CALB1, TRPV5 (low to none); Low to No CALCA and KIT in C-CD-IC-A
		CD-PC (general)	CD-PC	*AQP2*, AQP3*, FXYD4*, SCNN1G*, GATA3**	
		C-CD-PC	C-CD-PC		
		M-CD-PC	M-CD-PC		
		Outer medulla-CD-PC	OM-CD-PC		
		Inner Medulla-CD-PC	IM-CD-PC	*AQP2*, SLC14A2*	
		Transitional PC-IC	tPC-IC	*FXYD4*, SLC4A9*/SLC26A7**	
		CD-IC (general)	CD-IC	*CA2, ATP6VOD2**	
		CD-IC-A (general)	CD-IC-A	*SLC4A1, SLC26A7*, TMEM213**	
		C-CD-IC-A	C-CD-IC-A	*SLC26A7*, SLC4A1**	
		M-CD-IC-A	M-CD-IC-A	*SLC26A7*, SLC4A1, KIT*, CALCA*	
		CD-IC-B (general)	CD-IC-B	*SLC4A9*, SLC26A4**	
		C-CD-IC-B	C-CD-IC-B		
		M-CD-IC-B	M-CD-IC-B		
Vessels	Endothelial Cells (non glomerular)	Endothelial Cell (general)	EC	*EMCN*, PECAM1*, FLT1**	
		EC-Afferent/Efferent Arteriole	EC-AEA	*SERPINE2*, TM4SF1**	likely PALMD

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

		EC-Peritubular capillaries	EC-PTC	*PLVAP**	
		EC-Descending Vasa Recta	EC-DVR	*TM4SF1*, PALMD*	
		EC-Ascending Vasa Recta	EC-AVR	*DNASEIL3**	PALMD low to none
		EC-Lymphatics	EC-LYM	*MMRN1*, PROX1*	
Structure/Region	**Sub structure/Sub region**	**Cell Type**	**Abbreviation**	**Subset of Marker Genes**	**Pertinent negatives/comments**
Interstitium	Stroma (non glomerular)	Vascular Smooth Muscle/Pericyte (general)	vSMC/P	*TAGLN, ACTA2, MYH11, NTRK3, MCAM*	
		vSMC/P-Renin	vSMC/P-REN	*REN*	
		Fibroblast	FIB	*DCN, ZEB2, C7, LUM*	
	Immune	Macrophages-Resident	MAC-R	*CD163, IL7R*	
		Macrophage	MAC	*S100A9*	
		Natural Killer Cell	NKC	*NKG7*	
		Dendritic Cell	DC	*APOE*	
		Monocyte	MON	*C1QA, HLA-DRA*	
		T lymphocyte (general)	T	*CD3*	
		T Cytotoxic	T-CYT	*GZMA*	
		B lymphocyte	B	*IGJ*	
Structure/Region	Sub structure/Sub region	Cell Type	Abbreviation	*Marker Genes*	Pertinent negatives/comments
Interstitium	Stroma (non glomerular)	Vascular Smooth Muscle/Pericyte (general)	vSMC/P	*TAGLN, ACTA2, MYH11, NTRK3, MCAM*	
		vSMC/P-Renin	vSMC/P-REN	*REN*	
		Fibroblast	FIB	*DCN, ZEB2, C7, LUM*	
	Immune	Macrophages-Resident	MAC-R	*CD163, IL7R*	
		Macrophage	MAC	*S100A9*	
		Natural Killer Cell	NKC	*NKG7*	
		Dendritic Cell	DC	*APOE*	
		Monocyte	MON	*C1QA, HLA-DRA*	
		T lymphocyte (general)	T	*CD3*	
		T Cytotoxic	T-CYT	*GZMA*	
		B lymphocyte	B	*IGJ*	

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 6: Challenges and opportunities ahead

Challenges	**Explanations**
KPMP policies to address changes in technologies	Some technologies and platforms may change. How will data be acquired and archived to be compatible with data from future platforms?
Software changes (analytic or visual software)	New software present challenges in compatibility, reliability and security.
Strategies to test batch effect and technology drifts	Current and future technologies are expected to provide an *a priori* plan to detect batch effect and technology drifts and provide solutions. What reference tissue standards are suitable for this purpose?
Validation of emerging technologies and incorporation into KPMP.	Will there be a need for a standard tissue used for validating new technologies? What should this tissue be? Is there unlimited supply?
Hyperdimensional data management, storage and sharing	An increasingly problematic issue when big data will be generated from each tissue specimen.
Patient protection in the era of artificial intelligence.	The risk of linking patients to de-identified raw data may increase as machine learning tools develop further. Steps to mitigate risks in publicly available data will need to be implemented.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

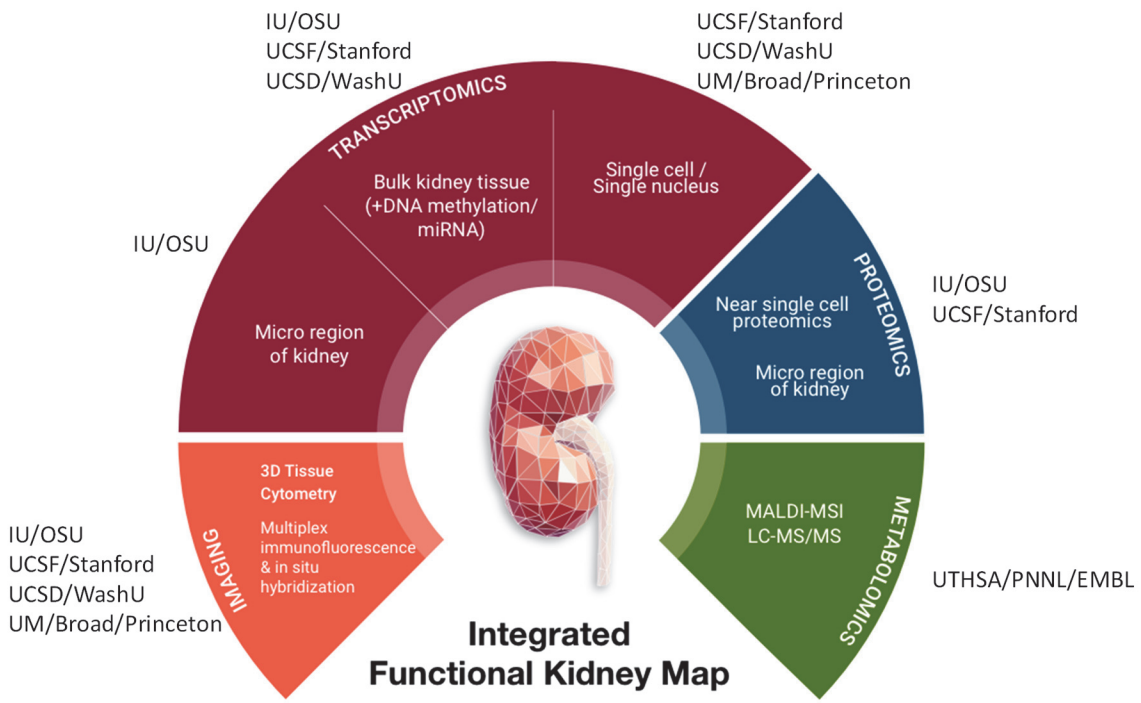


Fig. 1. Overview of Tissue Interrogation Sites and technologies in KPMP. IU- Indiana University; OSU – Ohio State University; UCSD – University of California, San Diego; WU – Washington University in St. Louis; UM – University of Michigan; UCSF – University of California, San Francisco; UTHSA – University of Texas Health Sciences, San Antonio; PNNL – Pacific Northwest National Laboratory; EMBL – European Molecular Biology Laboratory.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

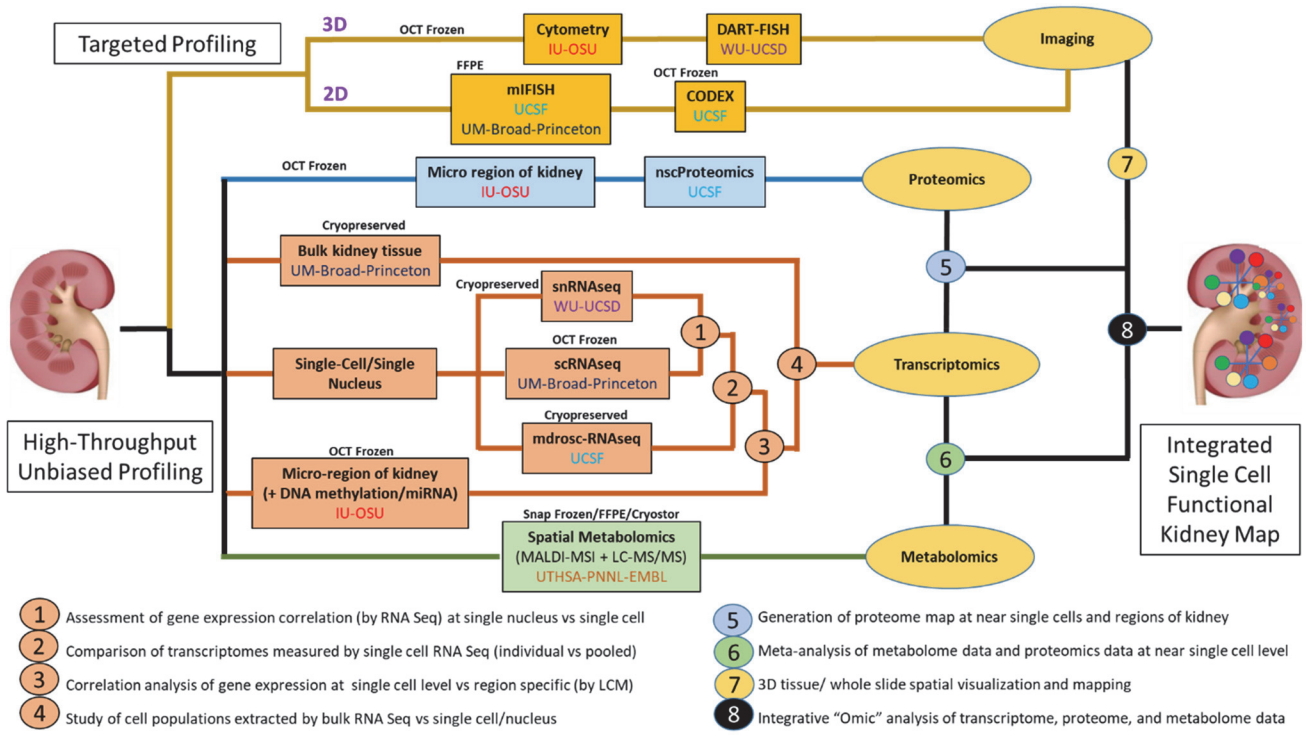


Fig. 2. The blueprint for molecular assessment in KPMP. Envisioned pathway template of detailed tasks and steps of the processes leading to the kidney atlas.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

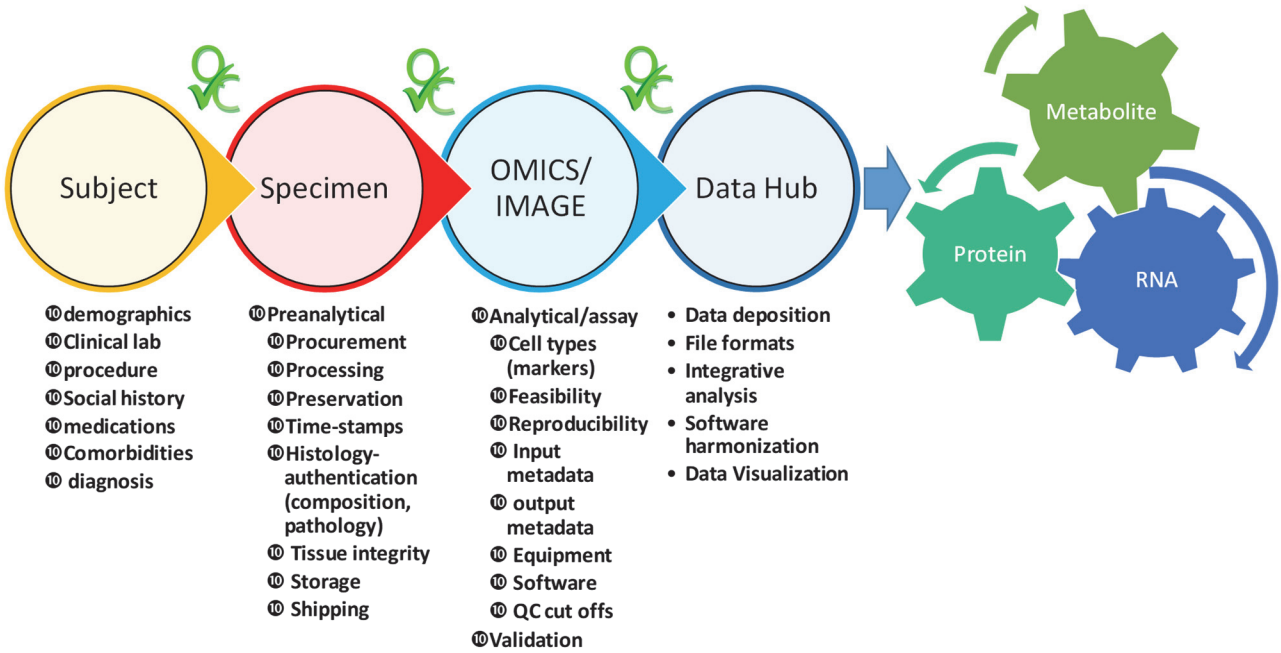


Fig. 3. Overview of "Follow the Tissue" pipeline with essential QA/QC parameters.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

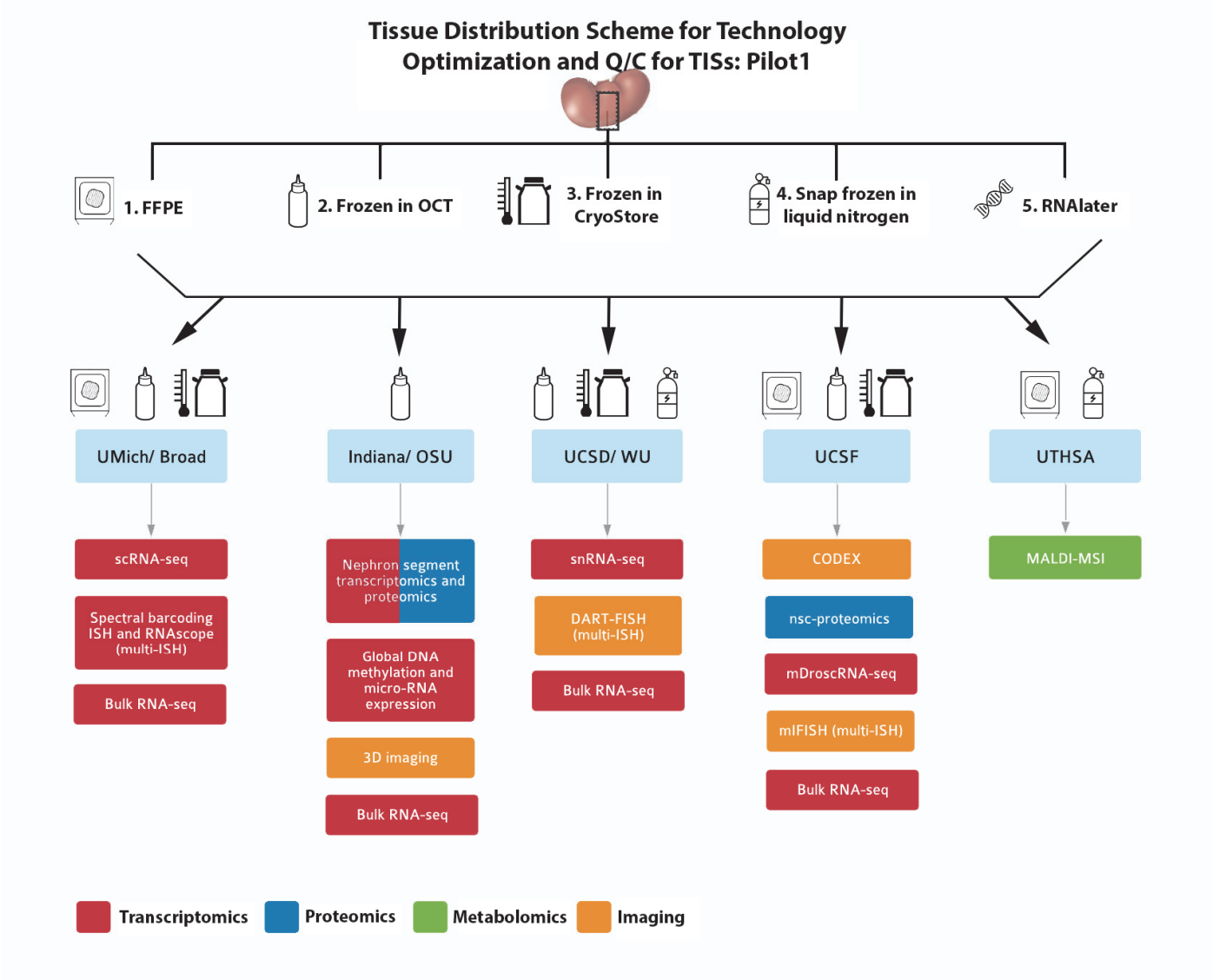
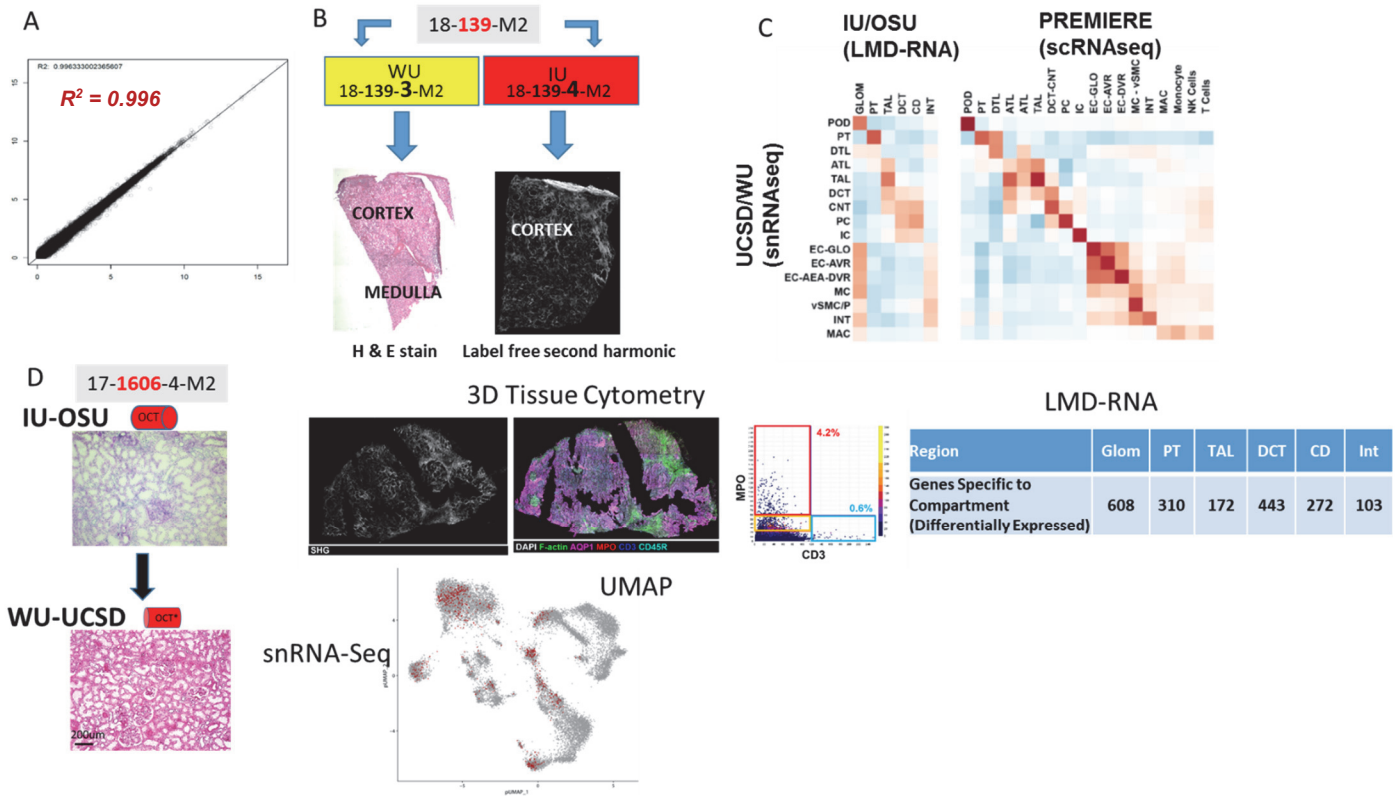


Fig. 4. Pilot 1 tissue processing and distribution. A single source kidney tissue from a tumor nephrectomy was sectioned into equal blocks (schema in supplemental figure 1) and processed into 5 tissue processing methods according to the downstream applications listed. Blocks of the tissue was distributed to the sites according to the technology as indicated by the drawings. Formalin fixed paraffin embedded (FFPE); in-situ hybridization (ISH); single cell (sc) and single nuclear (sn) RNA sequencing (RNA-seq). Near single cell (nsc)-proteomics; matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI MSI).

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Region	Glom	PT	TAL	DCT	CD	Int
Genes Specific to Compartment (Differentially Expressed)	608	310	172	443	272	103

Fig. 5. Testing the "Follow the tissue" pipeline with data generated from pilot samples and demonstrating feasibility of multiple technologies on limited tissue sample across sites. **A)** High correlation of bulk RNA seq from adjacent samples shipped and processed at different sites. **B)** Adjacent samples from same patient shipped to different TISs show different types of information from different technologies in KPMP with histological tissue composition verification for snRNAseq at WU-UCSD and label-free second harmonic generation highlighting extent of fibrosis at IU-OSU. **C)** Correlation between cell types and regions among snRNA-seq, LMD transcriptomics of the indicated regions by IU/OSU and between snRNA-seq and scRNA-seq. **D)** Economizing tissue usage by rotating fresh OCT frozen kidney tissue between IU and WU/UCSD. Here frozen sections were processed and analyzed for 3D tissue cytometry and LMD-RNA analysis and then the block was shipped to WU/UCSD for histological registration and subsequently generating snRNA-seq. Genes specific to each LMD compartment were determined by their upregulation in that compartment compared to other compartments as well as differential expression at a nominal $p < 0.001$. The UMAP plot shows the sample (~900 nuclei) contributing to many of the cortical cell types in the kidney snRNAseq atlas (~18000 nuclei) (13).

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

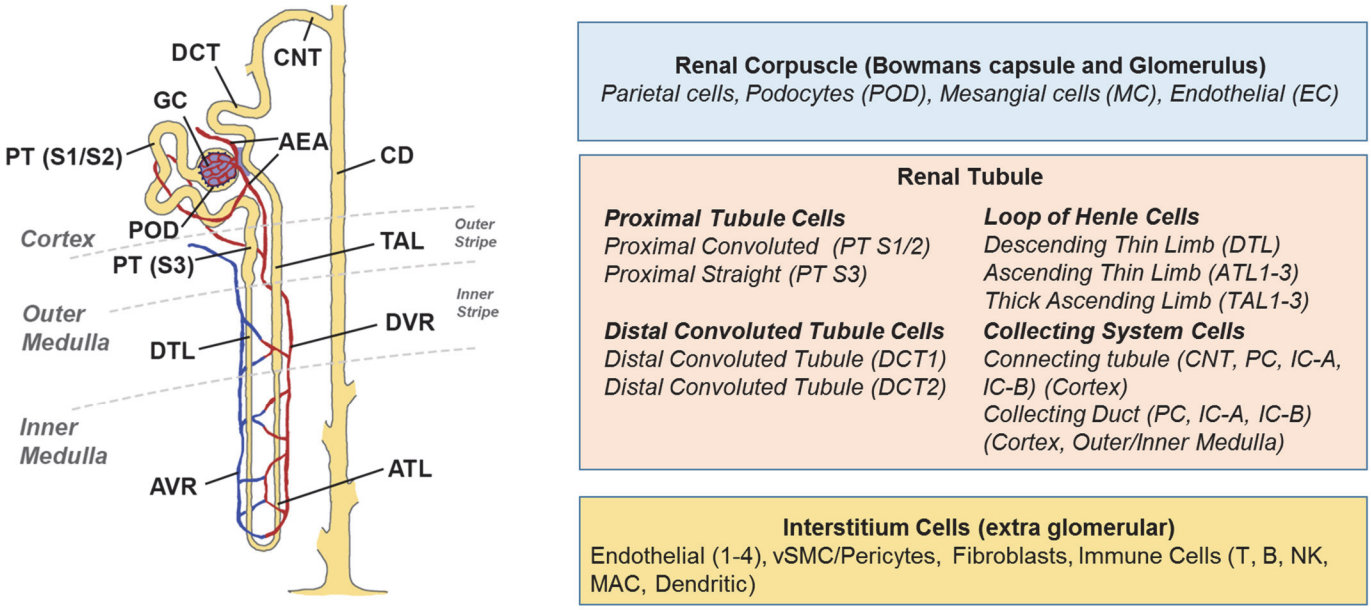


Fig. 6. Kidney regions and cell types. The nephron illustration is based on Kriz, W. & Bankir, L. A standard nomenclature for structures of the kidney. The Renal Commission of the International Union of Physiological Sciences(IUPS).Kidney Int.33,1–7 (1988). T, T lymphocyte; B, B lymphocyte; NK, natural killer cell; MAC, macrophage.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

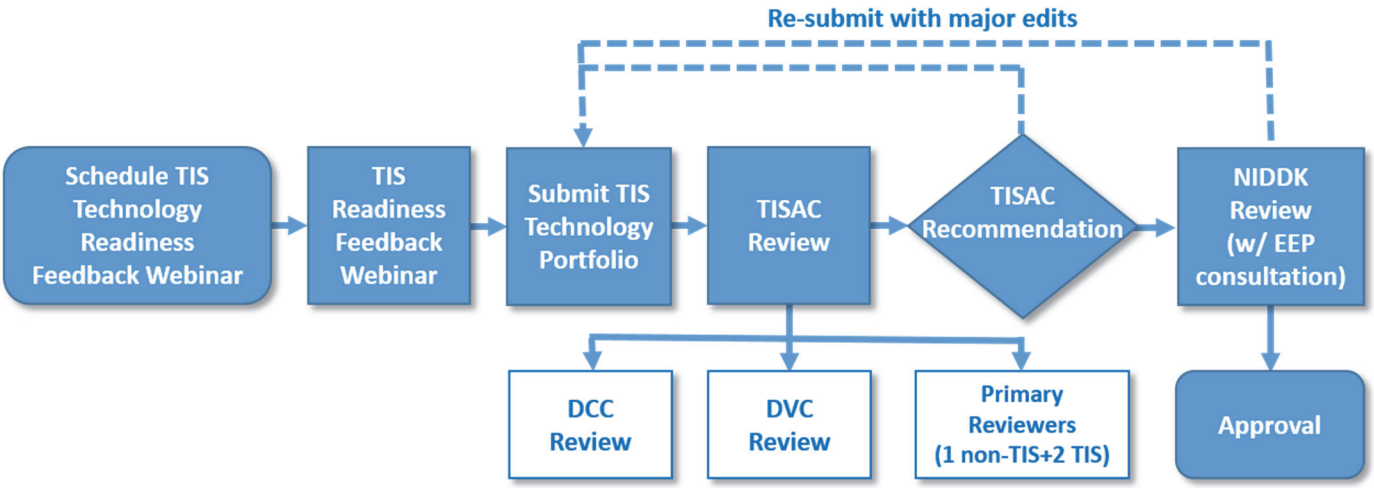


Fig. 7 The tissue interrogation site approval committee (TISAC) workflow. This workflow includes completion of a recorded webinar and submission of a portfolio that details the quality control metrics and technology outcomes. The TISAC reviews this material, providing recommendations to the tissue interrogation site for improvement before a site is approved to receive biopsies. The TISAC ultimately forwards their recommendations to the NIDDK and external expert panel (EEP) for final approval.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Hood L, and Rowen L. The Human Genome Project: big science transforms biology and medicine. *Genome Medicine.* 2013;5(9):79.
2. Ledergor G, Weiner A, Zada M, Wang SY, Cohen YC, Gatt ME, et al. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat Med.* 2018;24(12):1867-76.
3. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. *eLife.* 2017;6:e27041.
4. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med.* 2019;199(12):1517-36.
5. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nature Genetics.* 2014;46:543.
6. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. *Science.* 2017;356(6340):eaal3321.
7. Shema E, Bernstein BE, and Buenrostro JD. Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat Genet.* 2019;51(1):19-25.
8. Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science.* 2019;364(6441):685-9.
9. Mariani LH, Martini S, Barisoni L, Canetta PA, Troost JP, Hodgin JB, et al. Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies. *Nephrol Dial Transplant.* 2018;33(2):310-8.
10. Saran R, Robinson B, Abbott KC, Agodoa LYC, Bragg-Gresham J, Balkrishnan R, et al. US Renal Data System 2018 Annual Data Report: Epidemiology of Kidney Disease in the United States. *Am J Kidney Dis.* 2019;73(3S1):A7-A8.
11. All of Us Research Program I, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" Research Program. *N Engl J Med.* 2019;381(7):668-76.
12. Hu BC. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature.* 2019;574(7777):187-92.
13. Lake BB, Chen S, Hoshi M, Plongthongkum N, Salamon D, Knoten A, et al. A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat Commun.* 2019;10(1):2832-.
14. Menon R, Otto EA, Kokoruda A, Zhou J, Zhang Z, Yoon E, et al. Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development.* 2018;145(16).
15. Harder JL, Menon R, Otto EA, Zhou J, Eddy S, Wys NL, et al. Organoid single cell profiling identifies a transcriptional signature of glomerular disease. *JCI Insight.* 2019;4(1).
16. Hato T, Zollman A, Plotkin Z, El-Achkar TM, Maier BF, Pay SL, et al. Endotoxin Preconditioning Reprograms S1 Tubules and Macrophages to Protect the Kidney. *J Am Soc Nephrol.* 2018;29(1):104-17.
17. LaFavers KA, Macedo E, Garimella PS, Lima C, Khan S, Myslinski J, et al. Circulating uromodulin inhibits systemic oxidative stress by inactivating the TRPM2 channel. *Sci Transl Med.* 2019;11(512).
18. Micanovic R, Khan S, and El-Achkar TM. Immunofluorescence laser micro-dissection of specific nephron segments in the mouse kidney allows targeted downstream proteomic analysis. *Physiol Rep.* 2015;3(2).
19. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* 2016;44(8):3865-77.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. Rovin BH, and Klein JB. Proteomics and autoimmune kidney disease. *Clin Immunol.* 2015;161(1):23-30.
21. Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat Commun.* 2018;9(1):882.
22. Prentice BM, Caprioli RM, and Vuiblet V. Label-free molecular imaging of the kidney. *Kidney Int.* 2017;92(3):580-98.
23. Darshi M, Van Espen B, and Sharma K. Metabolomics in Diabetic Kidney Disease: Unraveling the Biochemistry of a Silent Killer. *Am J Nephrol.* 2016;44(2):92-103.
24. Palmer A, Phapale P, Chernyavsky I, Lavigne R, Fay D, Tarasov A, et al. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat Methods.* 2017;14(1):57-60.
25. Alexandrov T, Ovchinnikova K, Palmer A, Kovalev V, Tarasov A, Stuart L, et al. METASPACE: A community-populated knowledge base of spatial metabolomes in health and disease. *bioRxiv.* 2019:539478.
26. Vasquez JJ, Hussien R, Aguilar-Rodriguez B, Junger H, Dobi D, Henrich TJ, et al. Elucidating the Burden of HIV in Tissues Using Multiplexed Immunofluorescence and In Situ Hybridization: Methods for the Single-Cell Phenotypic Characterization of Cells Harboring HIV In Situ. *J Histochem Cytochem.* 2018;66(6):427-46.
27. Winfree S, Ferkowicz MJ, Dagher PC, Kelly KJ, Eadon MT, Sutton TA, et al. Large-scale 3-dimensional quantitative imaging of tissues: state-of-the-art and translational implications. *Transl Res.* 2017;189:1-12.
28. Winfree S, Khan S, Micanovic R, Eadon MT, Kelly KJ, Sutton TA, et al. Quantitative Three-Dimensional Tissue Cytometry to Study Kidney Tissue and Resident Immune Cells. *J Am Soc Nephrol.* 2017;28(7):2108-18.
29. Winfree S, Dagher PC, Dunn KW, Eadon MT, Ferkowicz M, Barwinska D, et al. Quantitative Large-Scale Three-Dimensional Imaging of Human Kidney Biopsies: A Bridge to Precision Medicine in Kidney Disease. *Nephron.* 2018;140(2):134-9.
30. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell.* 2018;174(4):968-81 e15.
31. He Y SB, Ong E, Mariani L, Lienczewski C, Balis U, Kretzler M, Himmelfarb J, Bertram JF, Azeloglu E, Iyengar R, Hoshizaki D, Mooney SD. *International Conference on Biomedical Ontology 2018 (ICBO-2018),*. Corvallis, Oregon, USA.; 2018.
32. Ong E HY. *The 8th International Workshop on Vaccine and Drug Ontology Studies (VDOS-2019), in association with The 10th International Conference on Biomedical Ontology (ICBO-2019*. Buffalo, NY, USA; 2019.
33. Brunskill EW, and Potter SS. Gene expression programs of mouse endothelial cells in kidney development and disease. *PLoS One.* 2010;5(8):e12034.
34. Chen L, Lee JW, Chou CL, Nair AV, Battistone MA, Paunescu TG, et al. Transcriptomes of major renal collecting duct cell types in mouse identified by single-cell RNA-seq. *Proc Natl Acad Sci U S A.* 2017;114(46):E9989-E98.
35. Fetting JL, Guay JA, Karolak MJ, Iozzo RV, Adams DC, Maridas DE, et al. FOXD1 promotes nephron progenitor differentiation by repressing decorin in the embryonic kidney. *Development.* 2014;141(1):17-27.
36. Lee JW, Chou CL, and Knepper MA. Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes. *J Am Soc Nephrol.* 2015;26(11):2669-77.
37. Liu J, Krautzberger AM, Sui SH, Hofmann OM, Chen Y, Baetscher M, et al. Cell-specific translational profiling in acute kidney injury. *J Clin Invest.* 2014;124(3):1242-54.

bioRxiv preprint doi: https://doi.org/10.1101/828665; this version posted November 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

38. Rogers NM, Ferenbach DA, Isenberg JS, Thomson AW, and Hughes J. Dendritic cells and macrophages in the kidney: a spectrum of good and evil. *Nat Rev Nephrol.* 2014;10(11):625-43.
39. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
40. Volkert G, Jahn A, Dinkel C, Fahlbusch F, Zurn C, Hilgers KF, et al. Contribution of the alpha8 integrin chain to the expression of extracellular matrix components. *Cell Commun Adhes.* 2014;21(2):89-98.
41. Zeisberg M, and Kalluri R. Physiology of the Renal Interstitium. *Clin J Am Soc Nephrol.* 2015;10(10):1831-40.
42. Chen L, Clark JZ, Nelson JW, Kaissling B, Ellison DH, and Knepper MA. Renal-Tubule Epithelial Cell Nomenclature for Single-Cell RNA-Sequencing Studies. *J Am Soc Nephrol.* 2019;30(8):1358-64.