Developing a public data resource for the NCI's BPV program

Ping Guan and Helen M. Moore, for the BPV Program

National Cancer Institute, NIH, Bethesda MD 20892

ping.guan@nih.gov

Abstract

The Biospecimen Pre-analytical Variables (BPV) Program is a National Cancer Institute-sponsored study to assess the effects of pre-analytical factors on the molecular profile of biospecimens. A comprehensive set of molecular profiling data has been generated from experiments addressing cold ischemic time (delay to formalin fixation), time in formalin, freezing methods, and storage temperatures and durations.

The BPV data set is currently being prepared for submission to the NIH Database of Genotypes and Phenotypes (dbGaP) and is expected to be released by the end of 2018. These data will be available to the biomedical research community, via dbGaP's controlled access, to support the development of evidence-based best practices for fit-for-purpose collection, processing, and storage of biospecimens for cancer research.

Introduction

Despite the extensive use of FFPE specimens in research and clinical settings, more data is needed to understand how suboptimal or variable specimen handling practices affect molecular analysis. The BPV Program recruited cancer patients with 4 primary cancer types (kidney, colon, ovary and lung) who were consented to donate specimens. Surgical resection tissues and matched blood from eligible patients were collected and subjected to pre-defined preanalytical manipulations. Each specimen was annotated with 300+ data elements that cover steps in the collection, handling, and processing procedures, pathology review, and clinical information. RNA, DNA, protein, and metabolites from these biospecimens was analyzed on multiple molecular platforms to assess potential effects of these pre-analytical manipulations.

BPV Experimental Design

Patient Cohort

- Cancer patients undergoing surgical treatment for one of the following cancer types (primary tumors only):
 - Renal Carcinoma
 - Colorectal Adenocarcinoma
 - Lung carcinoma
 - Ovarian Carcinoma
- Informed consent for donating pre-anesthesia blood, surgical tissue, and all associated data.
- Patients had not received pretreatment such as chemotherapy, radiation therapy and/or immunotherapy for any previous or current cancer diagnosis.

Tissue Collection

Blood Collection

Time in Fixative		
Experimental Protocol	Delay Time to Fixation	Time in Fixative
A		6hrs
B		12hrs
C		23hrs
D		72hrs

Delay to Fixation		
Experimental Protocol	Delay to Fixation	Time in Fixative
E	1 hr	
F	2 hrs	
G	3 hrs	
H	12hrs	10-12hrs

Method of Snap Freezing and Storage		
Experimental Protocol	Snap Freezing method	Storage method
U	Dry ice	-80°C
V	Dry ice	LN ₂
W	LN ₂	-80°C
X	LN ₂	LN ₂

Storage Temperature and Time			
Storage Times	T0 (0-3 Mon)	T1 (6-8 Mon)	T2 (12 Mon)
Temp 1	-80°C	-80°C	-80°C
Temp 2	LN ₂	LN ₂	LN ₂

Preanalytical Conditions Evaluated

- Time delay to fixation (DTF)
- Time in fixative (TIF)
- Snap freezing method
- Storage temperature of frozen specimens
- Duration of storage of frozen specimens

BPV Study Data at dbGaP

dbGaP The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans (<https://www.ncbi.nlm.nih.gov/gap>).

dbGaP Submission Process

Sub-Studies	
BPV Cohort	Preanalytical Impacts on Global Metabolite Profiling
	Investigate the effect of the delay to fixation on the proteome and phosphoproteome
	Investigate the effect of storage conditions of tumor specimens on the proteome and phosphoproteome profiling
	Preanalytic Impacts on Genomic Sequencing by NGS technology
	Evaluation of frozen conditions on mRNA profiling by TaqMan assay
	mRNA signature for stratification by cold ischemia time
Preanalytical Impacts on Copy Number Variation Detection by aCGH Technology	

BPV Molecular Data and Analytical Platforms

Study Type	Cancer Type	Platform and Analysis
Copy Number Variation	Kidney	Genome-wide CNV evaluation by array-based genomic hybridization (aCGH)
Genotyping	Kidney, Colon, Ovary	Whole exome sequencing by NGS
RNA Expression	Kidney, Colon, Ovary	Transcriptome profiling by NGS
miRNA Expression	Kidney, Ovary	Human miRNA profiling by WaferGen SmartChip (Human miRNA Panel)
mRNA Expression	Kidney	Oncopanel evaluation by TaqMan OpenArray
mRNA Expression	Kidney	Cold-ischemic biomarker expression by TaqMan
Protein Expression (Tissue)	Kidney, Colon, Ovary	Total protein and phosphoprotein expression by MassSpec
Protein Expression (Plasma)	Kidney, Colon, Ovary, Lung	Global metabolite profiling by MassSpec

Other BPV Resources

Comprehensive Data Resource (CDR and CDR-Lite)

CDR is a web-based information management system for biospecimen collection:

Used for Biospecimen Preanalytical Variables (BPV) and Genotype-Tissue Expression (GTEx) Programs.

Code published at GitHub (<https://github.com/NCIP/CDR>)

CDR-Lite: a lightweight version of CDR that supports disease-centered specimen collection (tissue and blood).

Code published at GitHub (<https://github.com/NCIP/CDR-Lite>)

Controlled Vocabulary and Ontology for Biobanking

BPV Sample Access at Van Andel Research Institute

The remaining BPV specimens are currently managed through a collaboration between Van Andel Research Institute and NCI which evaluates requests and distributes the specimens for biospecimen research or clinical assay development. For more details, please visit <https://pbc.vai.org/bpv/>

Acknowledgements

We thank the research participants who generously donated their biospecimens. Biospecimen Source Sites and analytical partners included:

Boston Medical Center, Emory University, The University of New Mexico, The University of Pittsburgh, Leidos BioMedical Inc., Van Andel Research Institute, Advanced Cell Diagnostics (ACD), CBIIT at National Cancer Institute, Capron Inc., Georgetown University, Integrated BioBank of Luxembourg (IBBL), Metabolon Inc., MoCha Lab at Frederick National Lab for Cancer Research, and RBM-Myriad.