

## ICSB-2007: Bayesian Mixture Models for Phage Display Experiments: Translation from Mice to Humans

Kim-Anh Do<sup>1\*</sup>, Yuan Ji<sup>2</sup>, Guosheng Yin<sup>1</sup>, Mikhail Kolonin<sup>3</sup>, Jessica Sun<sup>3</sup>, Wadih Arap<sup>3</sup>, Renata Pasqualini<sup>3</sup>

1. Department of Biostatistics, U. T. M. D. Anderson Cancer Center, Houston, USA

2. Department of Bioinformatics and Computational Biology, U. T. M. D. Anderson Cancer Center, Houston, USA

3. Department of Genitourinary Medical Oncology, U. T. M. D. Anderson Cancer Center, Houston, USA

\*email: kim@mdanderson.org

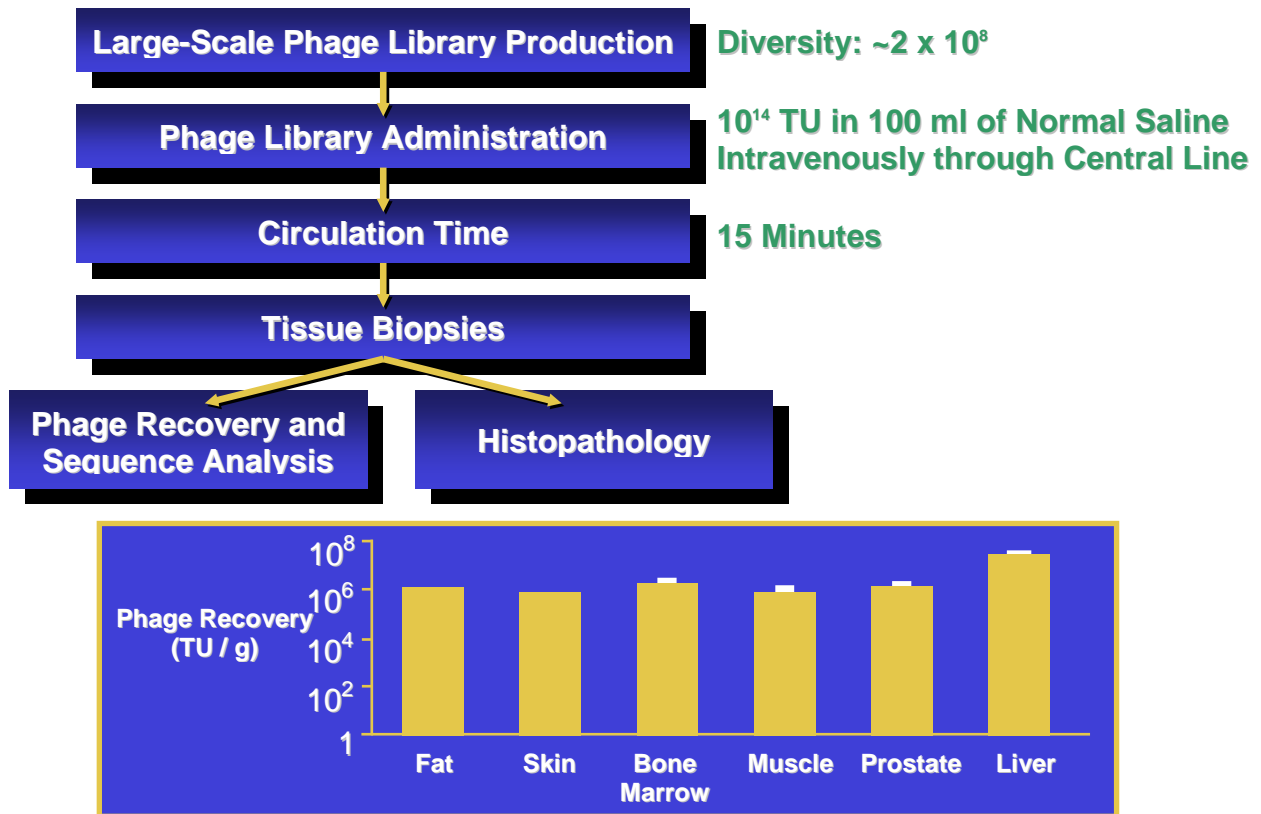
### Abstract

Phage display is a biological process to individually display up to millions of peptides and proteins on the surface of a small bacterial virus -- a phage. The use of phage display in screening for high-affinity ligands and their receptors has been crucial in functional genomics and proteomics. Our work was originally motivated by collaboration with molecular biologists who have been studying molecular targets that may be used to direct therapies to specific tissues. The long term translational goal of our research is that if drugs can be targeted to specific tissues in the body, then dosage can be altered to achieve the desired effect while minimizing side effects such as toxicity. Based on a count data set collected from six organs in mice: muscle, bowel, uterus, kidney, pancreas, brain, from an innovative multi-stage phage display experiment, we propose a class of Bayesian mixture models to cluster peptide counts into three groups that exhibit different display patterns across stages. Specifically, we propose a class of mixture models for analyzing the phage data with the following considerations. Since the recovered phage library at each stage was enriched before it was re-injected into the mouse for the next stage, the tri-peptide counts from the phage library were expected to increase across the multiple stages of the experiment if they were to bind to specific organs with high affinity. Likewise, the counts for the tri-peptides that did not bind would stay unchanged or even decrease, since only a certain number of peptides could bind to an organ. To capture these specific features in the data, we model the three counts of each peptide as three Poisson random variables and assume that the log Poisson means are expressed as a linear function of the stage index. Therefore, the sign of the slope in the linear function indicates whether the counts of the peptide decrease or increase over stages. For the prior distribution of the slope parameters, we propose a mixture of three normal distributions representing different trends in the display patterns of the peptides. Through posterior inference, we identify organ-specific tri-peptides that exhibit an ascending trend across consecutive stages. We apply a Bayesian false discovery rate approach to identify the peptides with the strongest affinity within the group with an ascending pattern. A list of peptides is obtained, among which important ones with meaningful functions are further validated by biologists. To examine the performance of the Bayesian model, we conduct a simulation study and desirable results are obtained.

Recent new data on an analogous experiment with human tissues (from brain-dead patients) are now available and work is in progress in developing a new Bayesian model based on Dirichlet process mixtures with utility functions to identify the enhanced tripeptides for specific tissues.

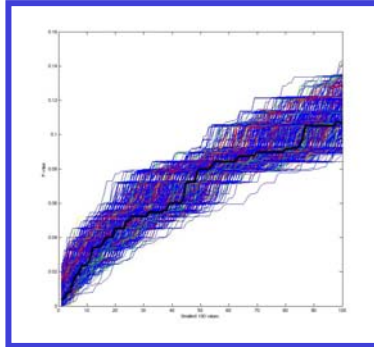
This research was supported in part by the University of Texas SPORE in Prostate Cancer grant CA90270.

**Figure 1: Strategy for Screening Phage Libraries in Humans**

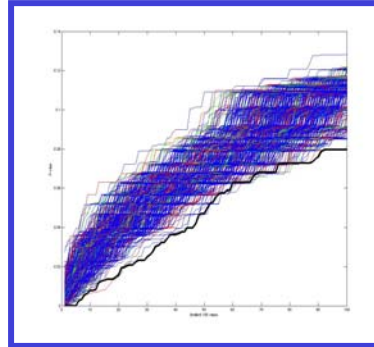


**Figure 2: Human Homing Motifs Compared to Library - Simulation Analysis based on permutation test**

**Round 1**



**Round 2**



**Round 3**

