Gene expression

Detecting and quantifying antibody reactivity in PhIP-Seq data with BEER

Athena Chen1, Kai Kammers2, H. Benjamin Larman3, Robert B. Scharpf2 and Ingo Ruczinski1,∗

1Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, 2Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA and 3Department of Pathology and the Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

∗To whom correspondence should be addressed.

Associate Editor: Christina Kendziorski

Received on February 24, 2022; revised on July 5, 2022; editorial decision on August 7, 2022; accepted on August 9, 2022

Abstract

Summary: Because of their high abundance, easy accessibility in peripheral blood, and relative stability ex vivo, antibodies serve as excellent records of environmental exposures and immune responses. Phage ImmunoPrecipitation Sequencing (PhIP-Seq) is the most efficient technique available for assessing antibody binding to hundreds of thousands of peptides at a cohort scale. PhIP-Seq is a high-throughput approach for assessing antibody reactivity to hundreds of thousands of candidate epitopes. Accurate detection of weakly reactive peptides is particularly important for characterizing the development and decline of antibody responses. Here, we present BEER (Bayesian Enrichment Estimation in R), a software package specifically developed for the quantification of peptide reactivity from PhIP-Seq experiments. BEER implements a hierarchical model and produces posterior probabilities for peptide reactivity and a fold change estimate to quantify the magnitude. BEER also offers functionality to infer peptide reactivity based on the edgeR package, though the improvement in speed is offset by slightly lower sensitivity compared to the Bayesian approach, specifically for weakly reactive peptides.

Availability and implementation: BEER is implemented in R and freely available from the Bioconductor repository at https://bioconductor.org/packages/release/bioc/html/beer.html.

Contact: ingo@jhu.edu

1 Introduction

Understanding interactions between the immune system and antigens is important to advance our understanding of the disease, disease progression and for the development of therapeutics. Phage ImmunoPrecipitation sequencing (PhIP-Seq) is a high-throughput technique to characterize antibody responses to extensive antigen libraries (Larman et al., 2011). Normalized sera are mixed with bacteriophages expressing the target peptides of interest, forming antibody–phage complexes (Fig. 1, left). These complexes are precipitated with magnetic beads, polymerase chain reaction (PCR) amplified and barcoded for multiplexing. The resulting product is sequenced and a matrix containing the number of sequencing reads with alignment to each peptide in the library for each sample is generated. Any alignment tool suitable for DNA or RNA short-read sequencing such as Bowtie (Langmead, 2010) or exact matching via grep can be used to align sequencing reads to the peptides in the library, and fastq files are then converted to read count matrices (for a detailed bioinformatics workflow including recommendations for demultiplexing and read alignment to the reference sequence database, see Mohan et al., 2018). PhIP-Seq libraries include VirScan (Xu et al., 2015), quantifying antibody binding to around 100 000 peptides spanning the genomes of more than 200 viruses that infect humans, among others. The primary objective is to identify antibody reactive peptides and to quantify the strength of antibody binding. PhIP-Seq experiments require the use of negative controls (so-called ‘beads-only’ or ‘mock immunoprecipitations’ lacking antibody input), which are included as 4–8 wells of a 96-well plate, and are compared to the read counts from each individual serum sample on the plate. We recently developed and implemented BEER, a Bayesian framework for identifying reactive peptides (Chen et al., 2022, preprint available at https://www.biorxiv.org/content/10.1101/2022.01.19.476926v1).

2 Features

PhIP-Seq experimental data are coordinated with sample and peptide metadata using the S4 class PhIPData (Chen et al., 2021). PhIPData extends SummarizedExperiment, enabling users to conveniently subset the data while preserving metadata information (Morgan et al., 2021). Peptide metadata can be added to a
2.2 Edger functionality

Since the output from both PhIP-Seq and RNA-Seq experiments are read count matrices, existing software for normalization and analysis of RNA-Seq data such as edgeR (Robinson et al., 2010) can also be applied to PhIP-Seq data. However, important differences in the data structures, experimental design and study objectives exist between the two approaches (Chen et al., 2022). We do recommend using existing functionality in the edgeR package for PhIP-Seq data to estimate the shape parameters for the Beta prior distributions for the probabilities of peptides pulling reads. edgeR uses an empirical Bayes approach to approximate the larger than binomial variability observed in the read counts, which improves performance compared to other estimation methods such as methods of moments and maximum likelihood estimation that do not borrow strength across peptides (Chen et al., 2022). Since edgeR does not rely on MCMC procedures and thus generates inference faster than BEER, we further investigated the performance of edgeR analyzing PhIP-Seq data in a two-group comparison with the mock IP samples in one group and a single serum sample in the other group (assuming equality of group variances) and found that BEER and edgeR are equally effective in detecting moderately and strongly reactive peptides, but BEER is needed to detect weakly reactive peptides (Chen et al., 2022). For user convenience, we included an edgeR-based pipeline for PhIP-Seq data in the BEER package. The method (classic likelihood ratio test or quasi-likelihood F-test) can be chosen using the \texttt{approx.method} argument in the \texttt{runBF} function. Two-sided edgeR \texttt{P}-values are converted into one-sided \texttt{P}-values (only enriched read counts indicate peptide reactivity) and are added with the estimated log fold changes to the \texttt{PhIPData} object via the function \texttt{runEdgeR}. Functionality also includes the round robin approach described above to assess the false-positive rate, which can be executed in \texttt{runEdgeR} with the argument \texttt{beadsRR}.

3 Conclusion

The Bioconductor package BEER provides two approaches for identifying reactive peptides from PhIP-Seq data, a Bayesian approach using an MCMC sampler, and an approach implemented in the RNA-Seq software package edgeR. The former is more sensitive detecting weakly reactive peptides, which may be particularly important when analyzing historic infections or the early stages of a developing antibody response. The implementations of the respective main functions \texttt{brew} and \texttt{runEdgeR} support parallelization for running multiple samples on one plate using the package BiocParallel in R (Morgan et al., 2022). BEER returns posterior probabilities for read count enrichment but does not explicitly generate a list of reactive peptides. We recommend to use posterior probabilities larger than 50% to delineate reactive peptides (Chen et al., 2022).

Funding

This work was supported by the United States National Institutes of General Medical Science NIGMS [R01 GM136724] and Allergy and Infectious Diseases [NIAID R01 AI095068], as well as the National Cancer Institute [NCI P50 CA062924 and NCI P30 CA006973].

Conflict of Interest: H.B.L. is an inventor on a patent describing the VirScan technology (US patent no. 15/105,722), a founder of Portal Bioscience, Alchemab and ImmuneID, and an advisor to TScan Therapeutics. R.B.S. is a founder and consultant of Deli Diagnostics, and owns Deli Diagnostics stock. Johns Hopkins University owns equity in Deli Diagnostics and Portal Bioscience.

Data availability

No new data were generated or analysed in support of this research.
References

Chen,A. et al. (2021) Container for PhIP-Seq Experiments. R package version 1.2.0. https://bioconductor.org/packages/release/bioc/html/PhIPData.html.
Chen,A. et al. (2022) Detecting antibody reactivities in phage immunoprecipitation sequencing data. bioRxiv: the preprint server for biology. https://www.biorxiv.org/content/10.1101/2022.01.19.476926v1.

Langmead,B. (2010) Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics, 32, 7.1–7.14. https://doi.org/10.1002/0471250953.bs1107s32.

Larman,H.B. et al. (2011) Autoantigen discovery with a synthetic human peptidome. Nat. Biotechnol., 29, 535–541.

Mohan,D. et al. (2018) Phip-seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. Nat. Protoc., 13, 1958–1978.

Morgan,M. et al. (2021) Summarized Experiment: Summarized Experiment Container. R package version 1.26.1. https://doi.org/10.18129/B9.bioc.SummarizedExperiment.
Morgan,M. et al. (2022) BiocParallel: Bioconductor Facilities for Parallel Evaluation. R package version 1.30.3. https://doi.org/10.18129/B9.bioc.BiocParallel.

Plummer,M. (2019) rjags: Bayesian Graphical Models using MCMC. R package version 4–10.

Robinson,M.D. et al. (2010) Edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26, 139–140.

Xu,G.J. et al. (2015) Viral immunology. comprehensive serological profiling of human populations using a synthetic human virome. Science, 348, aaa0698.