Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires

Jason J Lavinder[1,2], Andrew P Horton[3,4], George Georgiou[1,2,3,4,5] and Gregory C Ippolito[5]

Recent developments of high-throughput technologies are enabling the molecular-level analysis and bioinformatic mining of antibody-mediated (humoral) immunity in humans at an unprecedented level. These approaches explore either the sequence space of B-cell receptor repertoires using next-generation deep sequencing (BCR-seq), or the amino acid identities of antibody in blood using protein mass spectrometry (Ig-seq), or both. Generalizable principles about the molecular composition of the protective humoral immune response are being defined, and as such, the field could supersede traditional methods for the development of diagnostics, vaccines, and antibody therapeutics. Three key challenges remain and have driven recent advances: (1) incorporation of innovative techniques for paired BCR-seq to ascertain the complete antibody variable-domain VH:VL clonotype, (2) integration of proteomic Ig-seq with BCR-seq to reveal how the serum antibody repertoire compares with the antibody repertoire encoded by circulating B cells, and (3) a demand to link antibody sequence data to functional meaning (binding and protection).

Addresses
[1] Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78712-1062, USA
[2] Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712-1062, USA
[3] Center for Systems & Synthetic Biology, University of Texas at Austin, Austin, TX 78712-1062, USA
[4] Department of Biomedical Engineering, University of Texas at Austin, Austin, TX 78712-1062, USA
[5] Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712-1062, USA

Corresponding author: Ippolito, Gregory C (gci@mail.utexas.edu)

Current Opinion in Chemical Biology 2015, **24**:112–120

This review comes from a themed issue on **Omics**

Edited by **Benjamin F Cravatt** and **Thomas Kodadek**

http://dx.doi.org/10.1016/j.cbpa.2014.11.007

1367-5931/Published by Elsevier Ltd.

Introduction

Since the landmark discovery of antibody (or immunoglobulin, Ig) in blood serum more than 100 years ago, we now know conclusively that serum is composed of a complex spectrum of distinct antibodies (is polyclonal) and is generated by individual B-cell clones through extraordinary modes of genetic recombination, diversification, and selection by antigen (*anti*body *gen*erator) according to rules outlined in the paradigmatic 'clonal selection theory'. Remarkably, however, there had been no way to identify, and determine the relative concentrations, of the monoclonal antibodies (mAbs) that comprise the serum polyclonal pool elicited in response to vaccination or natural infection, until recently [1••,4••,5]. Understanding the composition of the antigen-specific serum antibody *protein* repertoire, the properties (e.g., affinities, epitopes recognized) of the respective Ig, and finally, the relationship between circulating Ig and the presence of clonally expanded peripheral B cells is profoundly important for the comprehensive understanding of humoral antibody responses.

The current era of modern genomics and proteomics is providing extraordinary new tools for examining antibody repertoires. Next Generation Sequencing (NGS) allows millions of B cell receptor (BCR) sequences to be obtained in a single experiment, and NGS approaches to studying the human antibody repertoire [6] not only aim to aid in the discovery of elite antibodies potentially useful as therapeutics, but also to comprehensively catalogue the antibody sequences that are elicited during an adaptive immune response [7]. Previously a limitation with NGS, the ability to obtain the endogenous variable heavy and light chain (VH:VL) pairs within NGS datasets is now feasible [1••,8••,9,10]. This paired VH:VL sequencing represents a major breakthrough in BCR repertoire analysis, obviating the need for multiplexed screening to identify functionally paired VH and VL. NGS has also provided a stepping stone to the direct characterization of serum antibodies (Ig) using NGS database-driven high-resolution protein mass spectrometry [1••,2,3,4••], providing a direct means to comprehensive delineation of the antibody repertoire.

Two antibody repertoires: the cellular and the serological

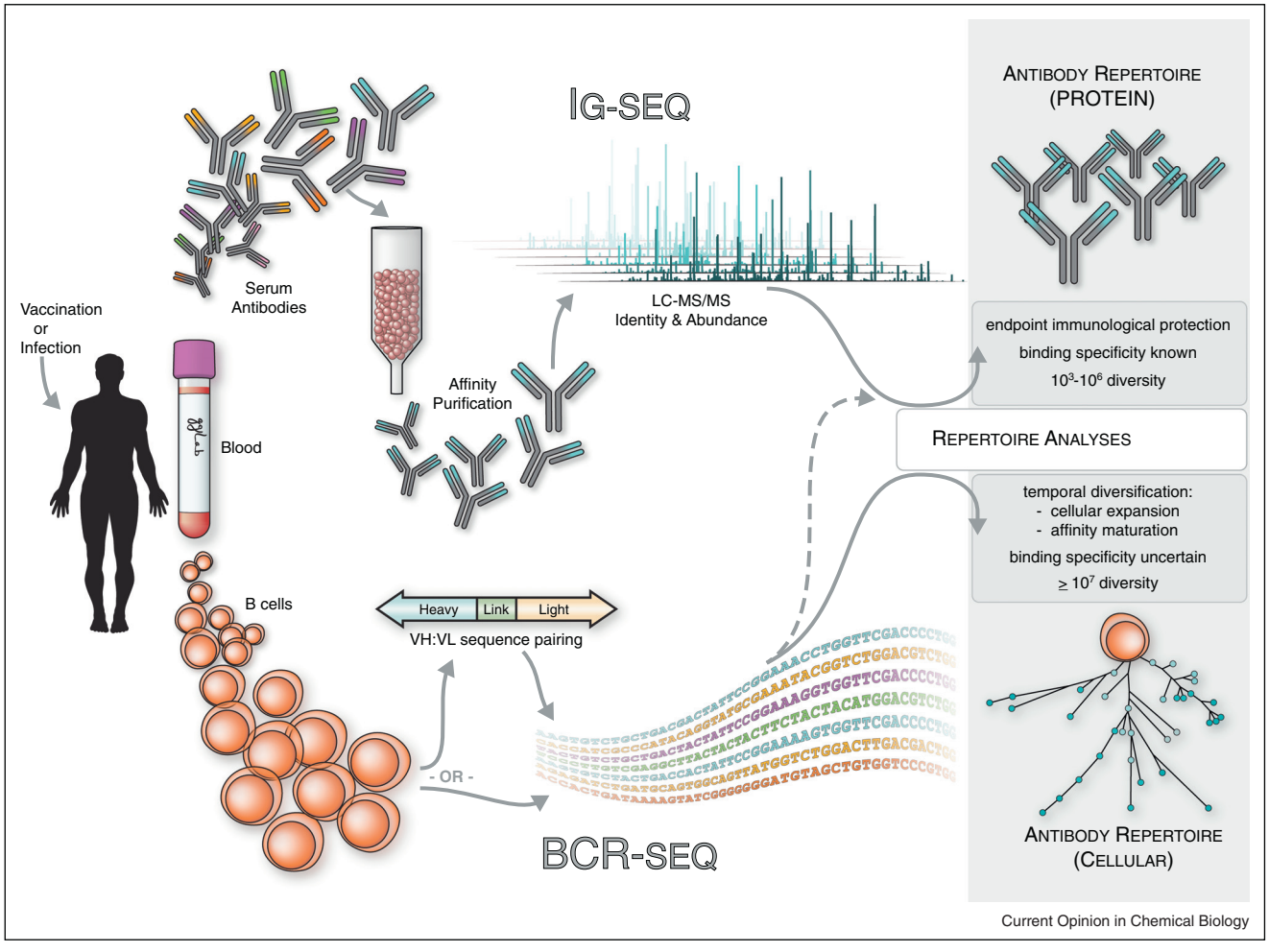
B cells, serum immunoglobulin, and persistence of the antibody repertoire

The mechanisms of antibody diversification and B-cell differentiation have been expertly reviewed elsewhere

[6,11,12] and are beyond the scope of this review. For purposes of introduction, it suffices to state that antibodies — the B cell antigen receptors (BCRs) — are composed of two heavy (H) and two light (L) polypeptide chains. Each chain consists of a constant (C) region and a variable region (VH or VL) which encodes the site of antigen binding. Antibody VH and VL domains each contain three juxtaposed spans of hypervariable sequence termed complementarity determining regions (CDRs), and their non-covalent association forms the antigen-binding site. CDR-H3 lies at the center of the antigen binding site, is the most diverse in terms of sequence, length and structures, and is typically a primary determinant of antibody specificity.

Whereas all newly formed B cells express antibody on their surface as BCR, and subsequently emigrate from their generative site in bone marrow to seed the periphery, it is only a small minority that might ultimately differentiate during the course of an immune response to become memory B cells (mBCs) and an even smaller fraction that secrete their BCR as soluble antibody. In this regard, we can conceive of the functional antibody repertoire as consisting of two major components: (1) the set of BCRs expressed on the surface of B lymphocytes, and (2) the collection of soluble Ig found in blood and secretions, produced predominantly (>95%) by terminally differentiated plasma cells in the bone marrow (BMPCs) [13] (Figure 1). Humoral immunity against pathogens can be

Figure 1



Analysis of human antibody repertoires from peripheral blood. The functional antibody repertoire consist of two major components: (bottom) the total set of BCRs expressed on the surface of peripheral blood B cells, and (top) the collection of soluble serum antibody circulating in the blood. The ability to compare and functionally characterize these two types of antibody repertoires provides a new paradigm in the study of the humoral response. This involves the isolation and proteomic analysis of affinity purified serum antibody (Ig-seq, top) in parallel with VH:VL pairing and/or NGS of peripheral B cell V gene repertoires (BCR-seq, bottom). The bioinformatic analyses of both the diversified cellular humoral immune response and the endpoint serological antibody response provides an avenue for effective antibody discovery, exhaustive antibody repertoire characterization, and an improved understanding of humoral immunity.