Experimental and Molecular Pathology 94 (2013) 182–187

Contents lists available at SciVerse ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

PCR amplification and high throughput sequencing of immunoglobulin heavy chain genes from formalin-fixed paraffin-embedded human biopsies

Hilla Tabibian-Keissar [a,b,*,1], Ginette Schibby [b,1], Miri Michaeli [a,b], Aviya Rakovsky-Shapira [a,b], Noemie Azogui-Rosenthal [a,b], Deborah K. Dunn-Walters [c], Kinneret Rosenblatt [b], Ramit Mehr [a,2], Iris Barshack [b,d,2]

[a] The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel
[b] Department of Pathology, Sheba Medical Center, Ramat Gan, Israel
[c] King's College London Medical School, London, UK
[d] Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

ARTICLE INFO

Article history:
Received 2 August 2012
Available online 27 August 2012

Keywords:
Formalin-fixed paraffin-embedded tissues
High throughput sequencing
Ig variable region genes
PCR

ABSTRACT

The use of high throughput sequencing (HTS) technologies in biomedicine is expanding in a variety of fields in recent years. The 454 system is an HTS platform that is ideally suited to characterize B cell receptor (BCR) repertoires by sequencing of immunoglobulin (Ig) genes, as it is able to sequence stretches of several hundred nucleotides. Most studies that used this platform for antibody repertoire analyses have started from fresh or frozen tissues or peripheral blood samples, and rely on starting with optimal quality DNA. In this paper we demonstrate that BCR repertoire analysis can be done using DNA from formalin-fixed paraffin-embedded (FFPE) human tissue samples. The heterogeneity of BCR repertoires we obtained confirms the plausibility of HTS of DNA from FFPE specimens. The establishment of experimental protocols and computational tools that enable sequence data analysis from the low quality DNA of FFPE tissues is important for enabling research, as it would enable the use of the rich source of preserved samples in clinical biobanks and biopsy archives.

Introduction

Following antigenic stimulation, B cells undergo clonal expansion and somatic hypermutation, increasing the diversity originally created by gene rearrangement. DNA sequencing of immunoglobulin (Ig) variable region genes is a common technique that is used for evaluation of antibody structure and function in health, disease and malignancy. Analysis of Ig variable region gene sequences can also provide important information regarding B cell clonal diversification, selection, and the somatic hypermutation (SHM) mechanism (Bognar et al., 2005; Dunn-Walters et al., 2002; Zuckerman et al., 2010a, 2010b). The limited output of sequence data provided by the classic Sanger sequencing method provides only a glance into lymphocyte repertoires (Banerjee et al., 2002; Ottensmeier et al., 1998; Tabibian-Keissar et al., 2008). In the past few years, several high throughput sequencing (HTS) technologies that yield many more sequences than the Sanger method have been developed. These technologies are rapid and effective and the price per sequenced nucleotide is lower than that of the Sanger method (Metzker, 2010; Schweiger et al., 2011; Zhang et al., 2011). Currently, the 454 system is the only HTS platform that is able to provide sequences of 250–500 bp. Weinstein et al. performed the first antibody gene repertoire analysis in zebrafish using this platform (Weinstein et al., 2009). This approach of analyzing Ig gene sequence data has already been used in the field of B cell receptor (BCR) and antibody repertoires, and already provided new insights into repertoire diversity (Boyd et al., 2010; Wu et al., 2010), B cell aging (Ademokun et al., 2011), and has proven the possibility of tracking rare lymphoma clones (Boyd et al., 2009; Briney et al., 2012; Lindner et al., 2012).

Analysis of Ig variable region genes, where the hallmarks of diversity such as SHM and combinatorial variation are spread over approximately 350 bp, requires long sequencing reads. The above-mentioned studies and others (Campbell et al., 2008; Scheid et al., 2009) used DNA from fresh or frozen tissues or peripheral blood samples. Several studies demonstrated successful HTS of short DNA fragments obtained from formalin-fixed paraffin-embedded (FFPE) tissues using the Illumina platform (Schweiger et al., 2009; Wood et al., 2010). Recently, Borras et al. performed mutation analyses of long PCR products (409 bp) of KRAS gene obtained by the 454 platform using FFPE tissues (Borras et al., 2011), but the mutation load in those sequences was lower than that of BCR genes.

Abbreviations: FFPE tissues, formalin-fixed paraffin-embedded tissues; HTS, high throughput sequencing; Ig, immunoglobulin; MID, multiplex identifier; SHM, somatic hypermutation.

* Corresponding author at: Department of Pathology, Sheba Medical Center, Ramat Gan 52621, Israel. Fax: +972 3 5308093.
E-mail address: hillatk73@gmail.com (H. Tabibian-Keissar).

[1] These authors contributed equally to this study.
[2] These authors supervised this project equally.

0014-4800/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yexmp.2012.08.002

In this paper we describe the experimental protocols we developed for preparing DNA samples from FFPE tissues for HTS targeted to BCR repertoire and mutational analyses using computational tools. We validate their reliability and demonstrate their usefulness by analyzing BCR repertoires in archived reactive human lymph node (LN) samples. High BCR repertoire heterogeneity was observed in all reactive lymph nodes analyzed, which indicates that the protocols are valid and potentially useful.

Methods

Samples

First phase: four FFPE reactive LN tissues, which have been archived about 10 years ago, were selected from the biopsy archives of the pathology institute at the Sheba Medical Center. In a second phase, more recent (one year) two FFPE samples and one frozen tissue from two cases, and two additional FFPE LN tissues (also archived about 10 years ago), were used. All cases were revised histologically using H&E staining. Demographic data and diagnoses are summarized in Table 1.

DNA extraction

Between 5 and 15 paraffin sections were cut from paraffin blocks, and de-paraffinization was performed by heat as follows: 200 μl PCR grade water was added to eppendorf tubes with the paraffin sections and tubes were heated to 95 °C for 5 min. The paraffin ring was removed after a centrifugation in full speed for 30 s; water was discarded.

DNA was extracted from the tissues (FFPE or frozen) by QIAamp DNA mini columns (Qiagen) according to the manufacturer's protocols with a few small modifications as follows. Proteinase K treatment was performed overnight; if the tissue was not completely digested then additional 10 μl of proteinase K was added and samples were incubated in 56 °C for 2 more hours. After proteinase K treatment, samples were incubated in 95 °C for 1 h (proteinase K treatment with heating is done in order to break the protein-DNA cross-links). In cases where the amount of tissue was small, samples were incubated at −18 °C for 1 h during ethanol precipitation. DNA was dissolved in 40–80 μl PCR grade water. DNA concentration and quality were determined by spectrophotometer measurements (NanoDrop 1000, Thermo Scientific).

PCR amplification of Ig variable region genes

Ig heavy chain variable region genes were amplified by semi-nested PCR using a set of 6 forward primers in the framework 2 (FR2) region specific for each VH family, and 2 reverse primers in the J region (one for each round, Supplementary Fig. 1). We found that using FR1 forward primers yielded too little DNA, as such long stretches of DNA are rarely well preserved. The 5′ sequence of the primer (IDT, HPLC grade) contained an adaptor that is specific for the 454-FLX emulsion PCR, followed by a 10 bp multiplex identifier (MID) tag, and completed by a ~20-bp primer sequence specific for each of the 6 VH families (Supplementary Fig. 1). The first PCR round included 6 tubes (one tube for each of the 6 FR2 primers); 2 μl of the first PCR round was used as a template for another PCR round which also included 6 tubes. PCR was performed with the XP Thermal Cycler (BioER). Each reaction contained 10–100 ng of genomic DNA as a template, 0.6–1.2 mM dNTP (Fermentas), 0.4 μM primers, 1 mM $MgCl_2$ and 1–2 U of proofreader Taq polymerase (Finnzyme). According to the manufacturer, the polymerase error rate is of an order of magnitude of 10^{-7} bp; we have checked this by amplifying and sequencing the actin gene from one DNA sample, and found no mutations (data not shown). Each round of PCR contained an initial denaturation step at 98 °C for 3 min. Denaturation and extension were carried out at 94 °C for 60 s and 72 °C for 90 s throughout, respectively. The annealing temperature in the first 5 cycles was 52.5 °C for 60 s, followed by 5 cycles at 52 °C and 25 additional cycles at 51.5 °C. The second PCR round was performed by the same cycling program as in the first round. PCR products were separated by electrophoresis on 2% agarose gel containing ethidium bromide. Sharp bands of around 300 bp were cut out of the gel, and DNA was cleaned by QIAquick gel extraction columns (Qiagen) according to the manufacturer's protocol.

High throughput sequencing by the Roche 454 flex titanium instrument

Two distinct rounds of HTS were performed the 454 flex titanium instrument by Dyn Diagnostics, Israel. For each round, DNA concentration and the quality of cleaned PCR products were measured by

Table 1
Patient demographic data, diagnoses and numbers of sequences obtained in each stage of HTS sequence data cleaning.

HTS round	1				2				
Sample/year	#1/1999	#2/1999	#3/1999	#4/1999	#5/2011		#6/2011	#7/1999	#8/1999
Type	FFPE	FFPE	FFPE	FFPE	Frozen	FFPE	FFPE	FFPE	FFPE
Sex	F	F	M	M	F		F	M	F
Age	54	44	48	63	57		50	78	32
Site of LN[a]	RT axillary LN	LT axillary LN	LT axillary LN	LT axillary LN	RT inguinal LN		LT inguinal LN	Right tonsil	Cervical LN
Diagnosis[b]	Axillary LA	Axillary LA	Axillary LA	Fever of unknown origin, LA	LA		LA	Tonsillar lymphoid area with acute inflammation and surrounding soft tissue	Florid reactive follicular hyperplasia
Sequences w/ identifiable MID tags	624	936	1437	3323	7847	2193	2397	1601	986
Sequences w/ identifiable primers	598	921	1423	3085	7758	2153	2334	1547	946
Unique sequences	259	270	553	402	5658	1119	1490	1014	640
No. VDJ combinations	50	49	35	68	907	239	260	46	22
No. clones\1000 unique sequences (rounded)	193	181	63	169	160	214	174	45	34

[a] RT = Right, LT = Left.
[b] LA = lymphadenopathy.