Reports

BioTechniques

Assigning immunoglobulin class from single-cell transcriptomes in IgA1-secreting versus membrane subpopulations

Colin Reily*,[1], Nuo Xu[2] & David K Crossman[3]

*[1]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [2]Department of Management, Information Systems & Quantitative Methods, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [3]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA; *Author for correspondence: creily@uab.edu*

BioTechniques 70: 89–99 (February 2021) 10.2144/btn-2020-0044

First draft submitted: 9 April 2020; Accepted for publication: 29 October 2020; Published online: 14 December 2020

ABSTRACT

IgA nephropathy (IgAN) is an autoimmune disease characterized by renal glomerular immunodeposits enriched for galactose-deficient IgA1 (Gd-IgA1; autoantigen) with the corresponding IgG autoantibodies. Despite the known contribution of Gd-IgA1 to IgAN, little is known concerning IgA1-secreting subpopulations responsible for autoantigen production. The goal of this study is to identify IgA1-secreting and membrane subpopulations from single-cell transcriptomic analysis. We developed a novel single-cell analytics workflow to discern cells expressing IgA1 secreted isoform or membrane-bound isoform. Multiple approaches were compared to assess immunoglobulin-isotype identity in single cells, and multiple immunoglobulin heavy-chain genes expressed in the same cells were found. To better identify specific immunoglobulin heavy-chain transcripts, we merged a software platform called Alteryx with the existing single-cell R toolkit program Seurat. This process allowed for improved calls on IgA1-secreting subpopulations based on secreting versus membrane splice-variant expression levels.

KEYWORDS:

Alteryx secreting ● bioinformatics ● IgA nephropathy ● immunoglobulin ● membrane ● scRNA-seq ● seurat

IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide [1]. This autoimmune disease is characterized by elevated blood levels of abnormally glycosylated IgA1 (galactose-deficient in some *O*-glycans; Gd-IgA1). Identifying IgA1-secreting subpopulations responsible for production of Gd-IgA1, and their unique transcriptional profile, would help identify critical mechanisms in the pathobiology of IgAN. To identify transcriptional mechanisms of Gd-IgA1 production in IgA1-secreting subpopulations, we used a previously generated biobank of Epstein–Barr virus (EBV)-immortalized B cells from peripheral blood of IgAN patients and healthy controls [2–6].

The aberrant glycosylation of Gd-IgA1 is related to abnormal expression and activity of key glycosyltransferases [2,3,5]. To assess mechanisms that drive abnormal glycosylation of IgA1 in cells from IgAN patients versus controls, we used high-throughput single-cell RNA sequencing (scRNA-seq) to map differential transcriptional responses in IgA1-secreting subpopulations. To identify IgA1-secreting cells, we used the *IGHA1* secreted splice variant. The splice variant is located on the 3′ end of the transcript, but the VDJ clonality – also of interest when discerning immunoglobulin function – is on the 5′ end of the gene. Thus we approached the analysis using scRNA-seq kits that targeted either the 3′ or the 5′ end (Figure 1). The purpose of this study was to develop a bioinformatic process to identify individual cells that have the IgA1-secreting isoform, the IgA1 membrane isoform or both isoforms; however, during our analysis of these IgA1 subpopulations, we found significant expression of other immunoglobulin heavy chains in the same cells, necessitating a process to identify which isotype class each cell should be called for.

This problem of needing sequence identification at both ends of the transcript is well recognized, and a recent publication has looked at using 3′ end reverse transcription with targeted capture and nanopore sequencing [7]. This study used both whole-transcriptome analysis from the 3′ end and long-read sequencing via Oxford Nanopore technology. The advantage was the availability of both whole-cell transcriptome and VDJ sequences; however, the process required significantly more PCR amplification upstream for hybridization capture, which can bias expression levels, and the nanopore sequencing technology generally has low recovery rates [7].

For our studies we needed an analysis process that could be hypothesis-driven, allowing for specific subpopulations determined *a priori*, and then we needed to perform both hypothesis testing and unbiased analysis on these subpopulations. For this, we turned to Alteryx, a software package used to handle large databases that allows users to generate unique workflows to quickly analyze data for gene expression profiles across all their datasets. Coupled with the R single-cell package 'Seurat', we generated the curated data sets from the EBV-immortalized cells derived from human peripheral blood B cells [8].

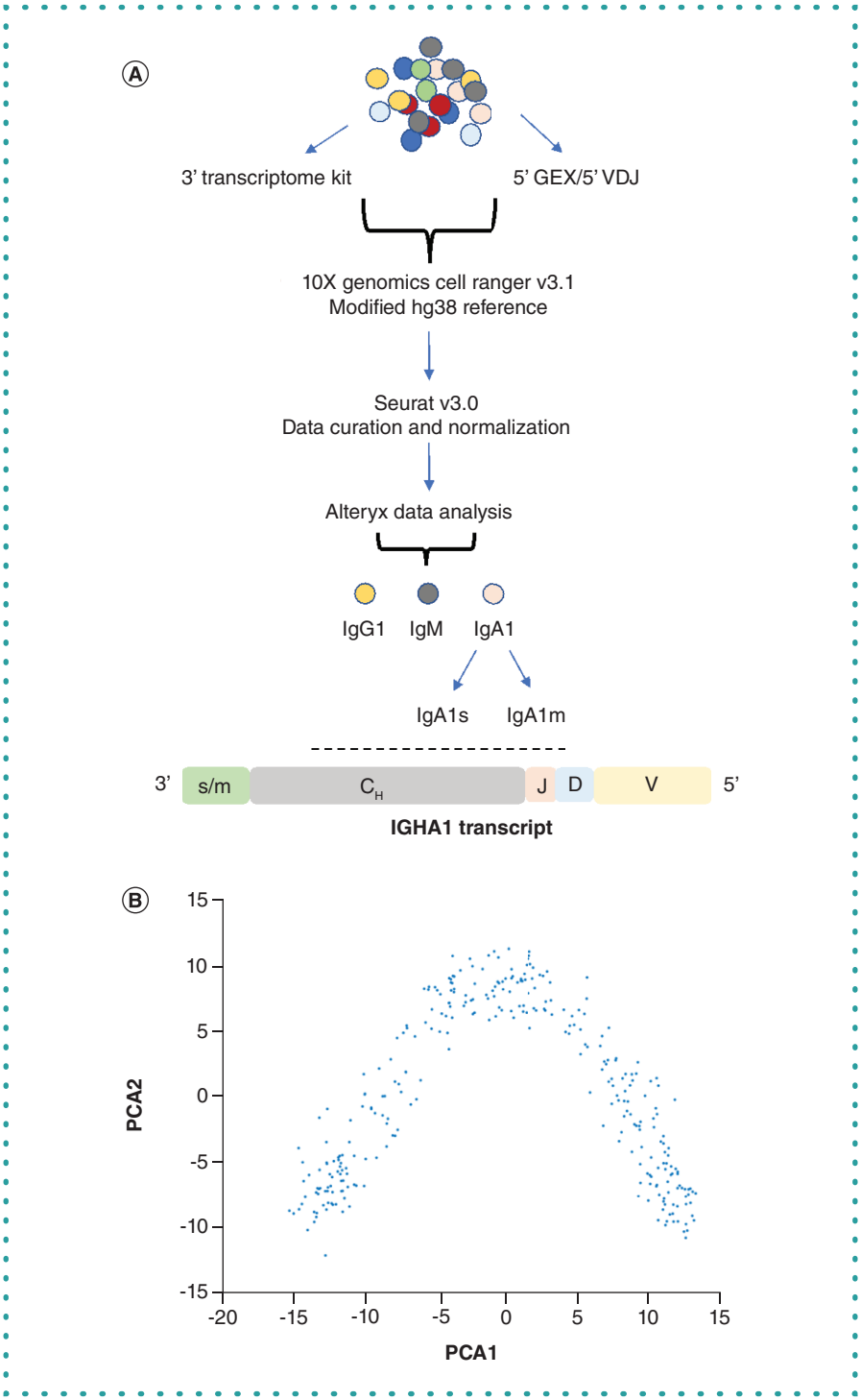

Figure 1. Overall scheme for data curation and analysis. (A) Multiple kits targeting the 3′ or 5′ end of mRNA transcripts were used. The 5′ VDJ kit was used as well to selectively amplify transcripts for sequencing the VDJ region. Sequencing data were aligned in Cell Ranger 3.1, curated and normalized in Seurat, then subgrouped in Alteryx. Cells were grouped by isotype heavy chain, followed by IGHA1 secretory or membrane form (s/m). The bottom diagram depicts regions of the IgA1 transcript, from the 3′ to the 5′ end, with the secretory and membrane splice variant on the 3′ end. **(B)** *IGHA1s*-expressing (IgA1s-secreting isoform) cells identified from the workflow were subject to PCA analysis.

Analysis of the raw and curated data from Seurat and Alteryx indicated that EBV-immortalized cells express multiple isotype classes of heavy chain genes; however, due to allelic exclusion during isotype switching, this is not biologically possible [9]. After analyzing the distribution profiles of expression of heavy chain isotypes in each cell, we defined an approach to make more appropriate calls of isotype heavy chains. With this process, we modified the hg38 reference database to include *IGHA1* splice variants for secretory and membrane-bound antibodies. This approach enabled a more accurate assessment of the immunoglobulin isotype calls and identification of critical IgA1-secreting subpopulations.

Materials & methods

A previously established biobank of EBV-immortalized peripheral blood mononuclear cells (PBMCs) was used [2,4–6]. Briefly, PBMCs isolated from patients with IgAN or other renal disease and healthy controls underwent EBV immortalization, a process that only targets B cells. For the purposes of this study, we only used IgAN donors, but immortalized B cells from healthy controls exhibit similar immunoglobulin heavy chain patterns [2]. Heterogenic mixtures (populations secreting multiple isotypes of immunoglobulin heavy chain) of B cells were grown in RPMI 1640 medium with 10% fetal bovine serum at 5% CO_2 [2,4,6]. Cells were centrifuged at 4°C, stored on ice for 30 min and, for the purpose of removing clumped and dead cells, were isolated as single live cells by using forward and side scatter in an Aria II flow cytometer before single-cell transcriptomic analysis. B cells were assessed using 10× Genomics 3′ transcriptome (v2.0, n = 4) and 5′ VDJ and GEX transcriptome kits (v1.1, n = 5), with target cell numbers of 3000 [10,11]. The target cell number of 3000 was used as recommended by 10× Genomics and See *et al.* who found that approximately 2000 cells were sufficient to target isotype subpopulations [12]. Transcripts were sequenced on an Illumina NextSeq 500, at 50,000 reads per cell. The sequences were called using 10× Genomics Cell Ranger v3.1. In addition to using the hg38 reference genome, splice variants for *IGHA1* secreted (ENSG00000282633.1) and membrane-bound (ENSG00000211895.5) (*IgHA1s* and *IgHA1m*) were added to the reference genome. Raw data were curated and normalized using Seurat v3.0 and analyzed using Alteryx Designer x64© in the following manner (Figure 1; example workflow in Supplementary Figure 1) [13]. Using the 5′ VDJ and GEX 10× Genomics kit, we identified antibody isotypes from the 5′ VDJ results and imported those barcodes into the 5′ GEX data for further comparison of immunoglobulin heavy chain calls.

Matrix data from Cell Ranger were read into Seurat v3.0 using the open source RStudio (v1.2.1335) [14,15]. Seurat was used to curate and normalize the data, using the instructions listed on the website (https://satijalab.org/seurat/v3.0/pbmc3k_tutorial.html):

- The Seurat object was created, and mitochondrial percentage assessed for each cell.
- Violin plots were created to visualize outliers and distribution characteristics. The EBV-immortalized cells are larger than primary B cells and have a higher mitochondrial content and gene set than primary cells (for visualization of primary PBMCs, see Vignettes in [14]). Thus for quality control (QC), minimum and maximum gene features were set at 1000/5000, and mitochondrial percentage at <20% (Figure 2 & Supplementary Figure 2).
- Data were then normalized using the LogNormalize function at 10,000.
- Normalized data from the Seurat object were copied into a new matrix and saved as an .rds file for future analysis. The following commands allow pulling out the data from the Seurat object and into an .rds file for further analysis by Alteryx:
    - normdataset <- pbmc[["RNA"]]@data
    - normnames <- normdataset@Dimnames
    - normdata <- normdataset@x
    - normmatrix <- matrix(normdata, ncol = "number of cells in object", nrow = "number of gene in object", dimnames = list(normnames[[1]], normnames[[2]]), byrow = TRUE)
    - normgenenames <- normnames [[1]].
- Both 'normmatrix' and 'normgenenames' were saved as .rds files.

To compare isotype class calls in Seurat and in our workflow, we identified a subset of *IGHA1s*-containing cells in Seurat using the following commands only after the Seurat object had been normalized and scaled using either the standard workflow or 'SCTransform' (for SCTransform we also regressed out mitochondrial percentage as suggested):

- IGHA1s_sub <- subset('SeuratObject', subset = IGHA1s >0)
- WhichCells(IGHA1s_sub)

The cell IDs were then imported into our Alteryx workflow and used to isolate those specific cells identified by Seurat as *IGHA1s* expressers. All immunoglobulin heavy chain isotypes for each cell were assessed.

To input the .rds files into Alteryx, we set up a workflow using the R script tool, with the following commands code:

- dat <- as.data.frame(readRDS("C:/filename/normmatrix.rds"))
- write.Alteryx(dat, 1)

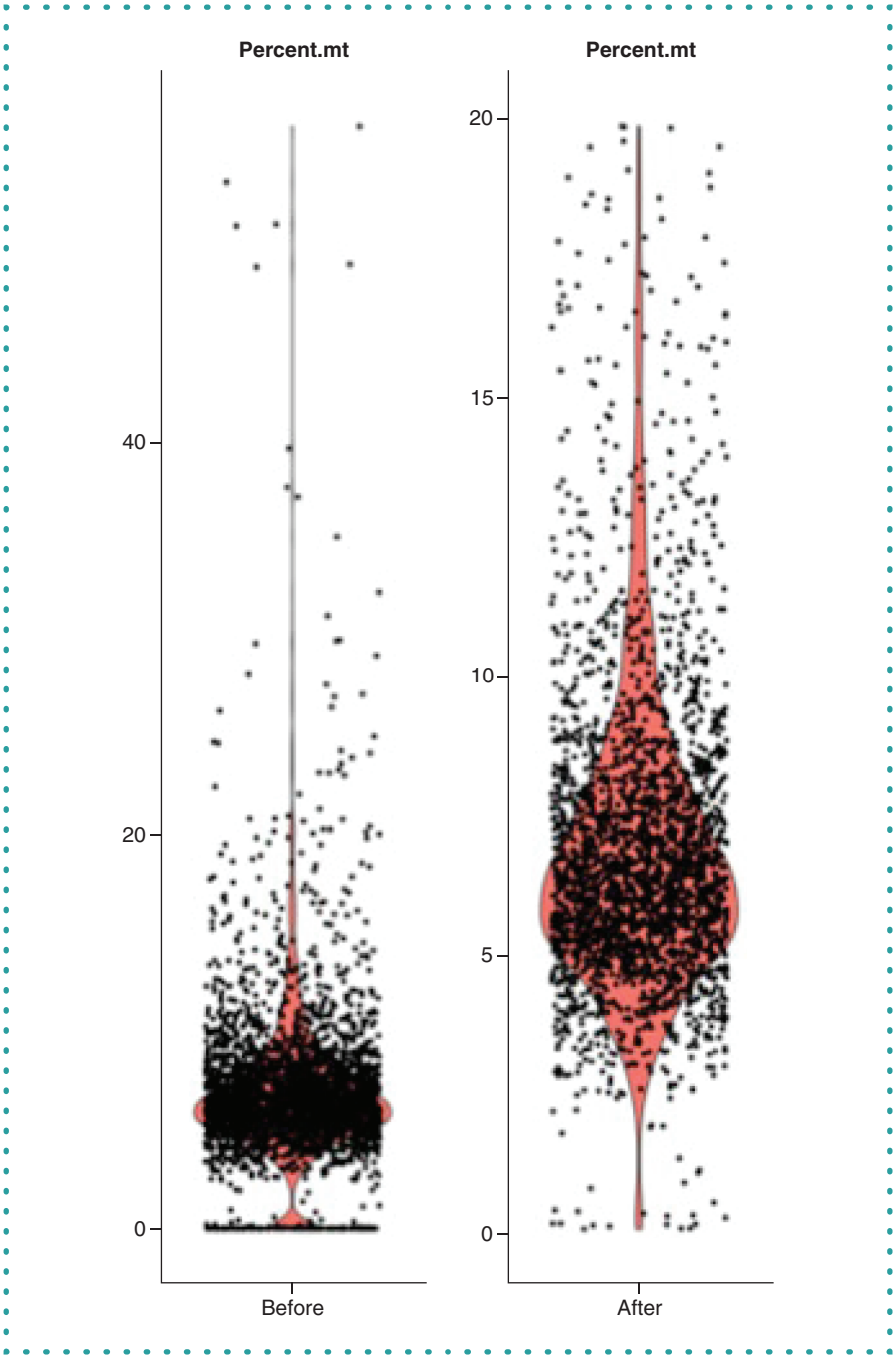

Figure 2. Mitochondrial percentage in EBV-immortalized cells before and after curation. Generally, these cells have higher expression of mitochondrial genes than their primary counterparts.

- dat2 <- as.data.frame(readRDS("C:/filename/normgenenames.rds"))
- write.Alteryx(dat2, 2)

Using the Output Data tool, we copied the new files in Alteryx database format (.yxdb) to the appropriate file.

Using Alteryx, we can query and categorize the data matrices from Seurat to find and analyze subpopulations of cells based on their gene expression profiles. This is done by creating workflows that employ a suite of tools designed to manage large datasets. Single-isotype heavy chain-expressing cells were subgrouped based on their expression isotype (*IGHM*, *IGHG1 etc.*), and *IGHA1* specifically into *IGHA1s* and *IGHA1m* subpopulations. Once specific subpopulations were grouped, either analysis was performed in Alteryx or the data were exported into a matrix table for further processing in R or other statistics packages.

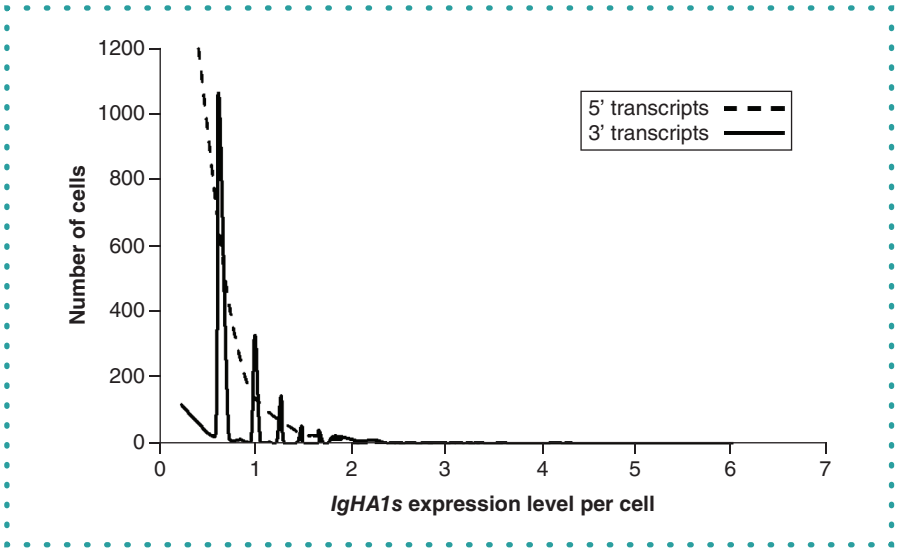

Figure 3. *IGHA1s* expression across all cells in two different samples. One sample of immortalized B cells was analyzed using a 5′ transcriptome kit (dashed line, 2326 cells), and the other using a 3′ transcriptome kit (solid line, 3233 cells). The Y-axis represents the number of cells that, based on the x-axis *IgHA1s* expression level, expressed a specific number of transcripts.

The relative homology of *IGHA1s* and *IGHG1* was assessed using cDNA transcripts from Ensembl (www.eseast.ensembl.org) of *IGHA1* secretory (ENST00000633714.1/ ENSG00000282633.1) and *IGHG1*-201/202/203 (ENST00000390542.6/ ENST00000390548.6/ ENST00000390549.6). Additional cDNA sequences used for comparison are: *IGHG2*-202 (ENST00000641095.1), *IGHG3*-202 (ENST00000641136.1), *IGHG4*-202 (ENST00000641978.1) and *IGHM*-202 (ENST00000637539.2).

Results & discussion

Immortalized B cells and IgA1-secreting cell lines derived from peripheral blood of patients with IgAN and healthy controls were used as model systems for mechanistic studies of Gd-IgA1 production [2]. EBV immortalization targets cells with CD21 receptors and thus creates a heterogenic mixture of B cells and immunoglobulin-secreting cells [16]. A mature B cell can be activated by a specific antigen and various cytokines to undergo a class switch from IgM/IgD to express other antibodies (IgA1/IgA2/IgG1-4/IgE). Once this occurs, the plasma cell or memory B cell does not produce any other antibody isotype [17]. With this in mind, we compared the expression of *IGHA1s* across all cells in two tested samples (Figure 3). As expected, we saw a range of expression levels, but we did not expect to see the *IGHA1s* gene expressed in all cells. As both samples were from a mixture of B cells that also secrete IgG and/or IgM (data not shown), we can assume that some or even a majority of the gene calls for *IGHA1s* were incorrect, and that most of the *IGHA1s* calls constituted background noise. These observations appear contradictory to the conventional doctrine of 'one B cell, one antibody'; however, another publication found similar results [18]. Using the same 10× Genomics 5′ VDG and GEX platform, Shi *et al.* found that a significant number of B cells (~10–40%) at various stages of differentiation express multiple immunoglobulin heavy chain transcripts [18]. Although the study did not assess the secretory sequence of the immunoglobulin heavy chain transcripts, the results were consistent with our findings.

To better elucidate the potential for confounding isotype heavy chain calls in the same cell, we compared *IGHA1s* expression across all cells with the ratio of other heavy chain isotypes (Figure 4). Figure 4A shows the relationship between *IGHA1s* and *IGHG1* expression in individual cells from a heterogenic B cell population assessed with a 5′ kit. This figure shows that, while there is some variability, as cells increase *IGHA1s* expression the relative amount of *IGHG1* also expressed appears stable. Figure 4B, a close-up panel of Figure 4A, highlights that the majority of cells expressing *IGHA1s* are low expressers and likely constitute background noise. Figure 4C shows the exact same cells as in Figure 4A & B, but includes information on expression of all other heavy chain isotypes for every cell. These data indicate that across all cells, there will be some level of expression of multiple heavy chain isotypes, confounding the ability to make an isotype-specific designation because the ratio of other isotype heavy chains is variable. This phenomenon is consistent across multiple cell lines, the only difference being the variability of contaminating heavy chain isotypes in different samples (Supplementary Figure 3A–C). We employed the 5′ VDJ-GEX kit from 10× Genomics, which sequences the VDJ transcripts in the same cells along with 5′ GEX transcriptomes. Using the Loupe Browser from 10× Genomics, we incorporated the 5′ VDJ data into the 5′ GEX transcriptome data and exported the cell barcode IDs specific to *IGHA1*-positive cell calls based on the VDJ analysis [19]. We used this subpopulation of cells to compare *IGHA1s* and *IGHG1* expression (Figure 5). Although the pattern of *IGHG1* expression to *IGHA1s* was slightly different (the cells in Figure 5 are the same cells as in Figure 4), we still found significant *IGHG1* expression across a significant number of moderate- to high-expressing *IGHA1s* cells using the 5′ VDJ scRNA-seq kit. While the power of this approach can yield sequence data on the VDJ

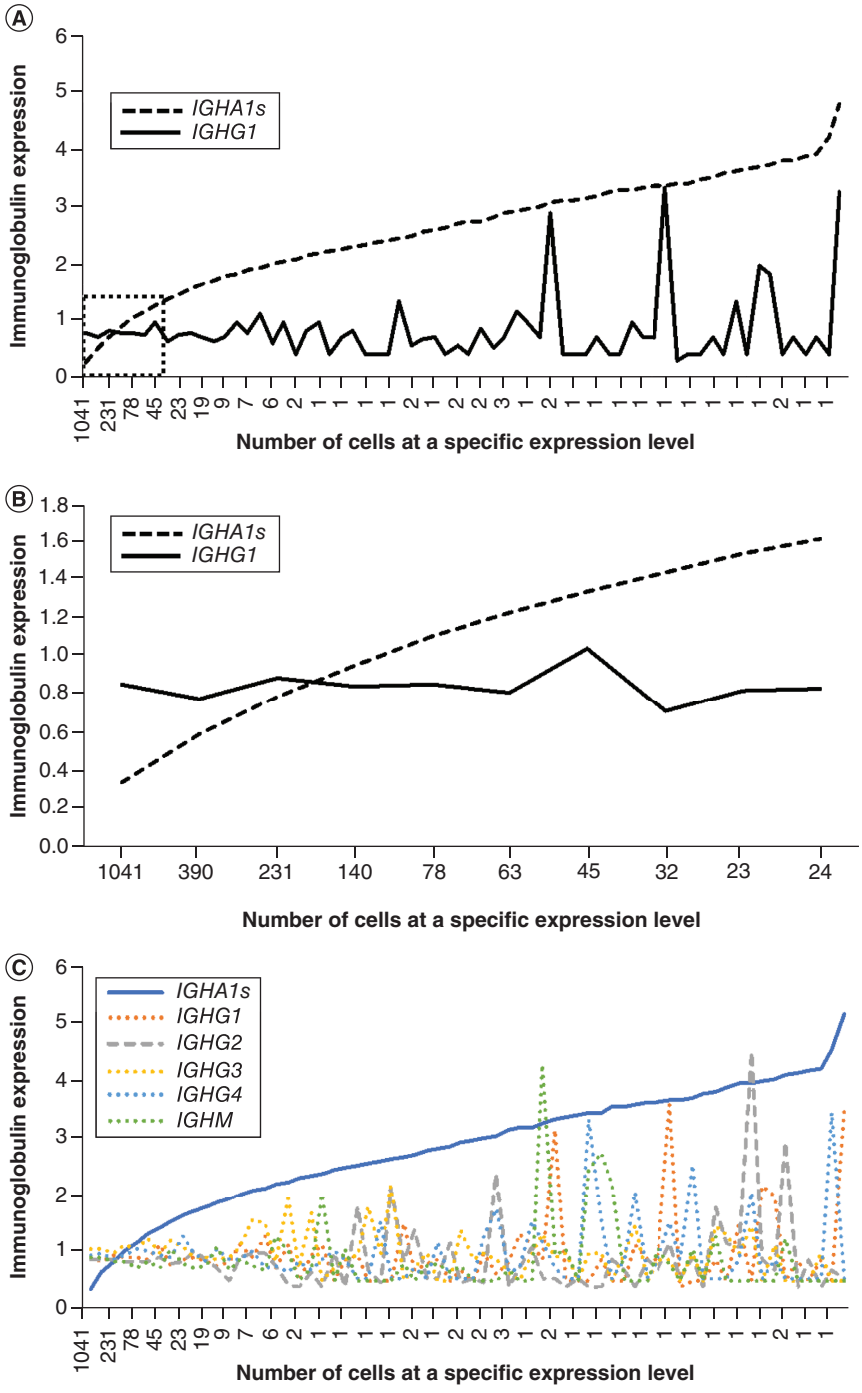

Figure 4. ***IGHA1s*** **and all other isotype heavy chain expression levels were assessed across all cells in a sample of immortalized B cells. (A)** ***IGHA1s*** **expression per cell compared with** ***IGHG1*** **expression for those same cells. The Y-axis represents the** ***IGHA1s*** **and** ***IGHG1*** **expression level. (B) Close-up of the bottom left panel of (A), showing that most of the cells in this heterogenic population have very low** ***IGHA1s*** **and** ***IGHG1*** **expression. (C)** ***IGHA1s*** **expression per cell compared with all other heavy chain isotype expression in those same cells. The Y-axis represents the** ***IGHA1s*** **expression level and the average** ***IGH*** **expression level for every other isotype. Numbers on the X-axis are the number of cells expressing** ***IGHA1s*** **at that specific level. Total cells = 2326.**

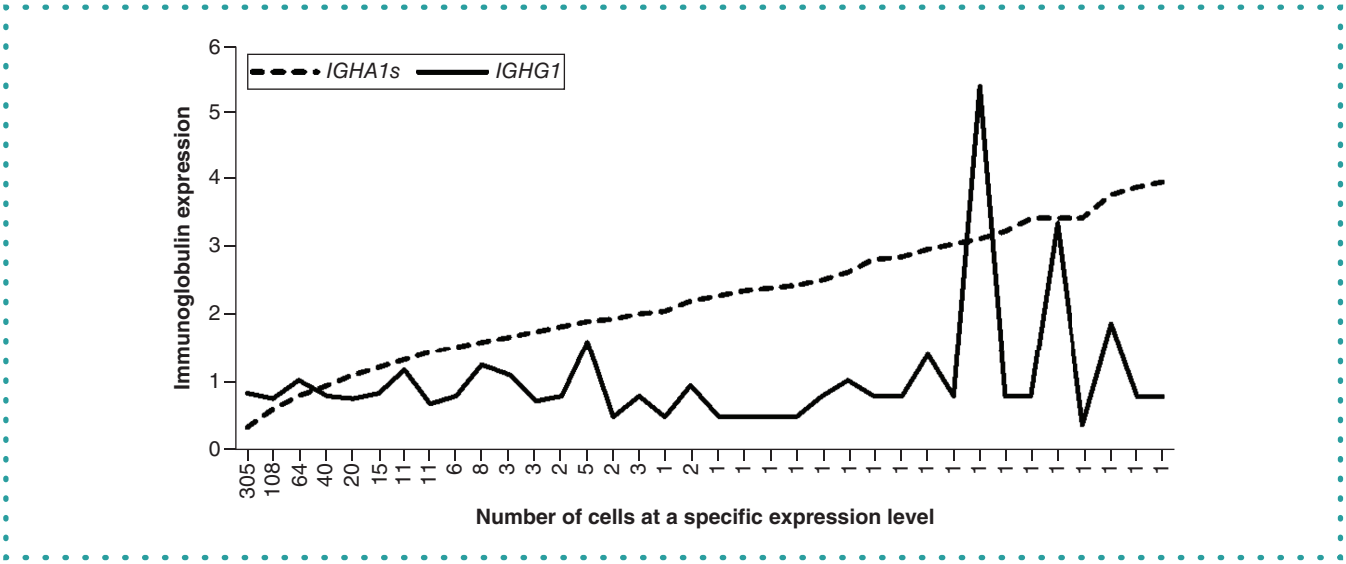

Figure 5. *IGHA1s* expression level compared with *IGHG1* expression level. Single-cell 5′ VDJ PCR was used to identify *IGHA1*-positive cells. These data were added to the 5′ GEX transcriptome in 10× Genomics Cell Loupe browser on 2218 immortalized B cells. *IGHA1s* and *IGHG1* expression levels were derived from the 5′ GEX data. The dotted line shows the *IGHA1s* expression level and the solid line shows the *IGHG1* expression level.

regions when combined with transcriptome profiling, the limitation is the poor coverage at the 3′ end where the secreted/membrane sequence is located.

Structural homology can be an issue in overlapping calls of similar sequences, but there is limited alignment similarity in the immunoglobin heavy chains [20]. Given that we were observing expression of multiple heavy chain isotypes in the same cell, the simplest answer is that we could be seeing doublets in our capture (i.e., two cells in one droplet), or another mRNA capture overlap from lysed cells. Therefore we checked whether the initial QC steps performed in Seurat to remove data outliers were sufficient. We compared our normal QC workflow with another one with a lower nFeature (3500) cutoff, a lower nFeature and nCount (15,000) and the new SCTransform function in Seurat (VlnPlots shown in Supplementary Figure 4). We found that these modified workflows did not affect the multiple isotype *IGH* calls in our cells (Supplementary Figure 5).

Given the overlap in calls among the different heavy chain isotypes, we needed to find the best route to assign the appropriate antibody isotype. A previous publication analyzing transcriptome and VDJ clonality in B cells from breast cancer tissue found similar issues with assigning isotype class and used a cut-off value of tenfold higher than the next highest expressing heavy chain isotype in that cell [21]. However, because each cell line likely has a unique proportion of immunoglobulin producers in the total population, using a predetermined expression cutoff would not work well. Using *IGHA1s*, we mapped the relative ratio of its expression to the next highest expressed heavy chain isotype (*IGH*) for each cell (Figure 6). We can calculate this relationship using the slope of the *IGHA1s*/*IGH*, and this can help determine whether the informatic calls for a positive *IGHA1s* cell are improving. It is important to note that the expression levels have been log-normalized, so biological expression differences are much larger. Although we found a significant positive correlation between cells with higher *IGHA1s* expression and an increased ratio of *IGHA1s*/*IGH* expression, we do not believe we should use a simple expression cut-off measurement to make appropriate isotype calls, because some cells with *IGHA1s* expression had low *IGHA1s*/*IGH* expression ratios (Figure 6). This phenomenon held true for three other cell lines assessed (Supplementary Figure 6).

Because we found consistent overlap in the heavy chain isotype calls in the same cell, we assessed expression across multiple heterogenic cell lines from various donors to determine what may be an appropriate ratio to make for a cell heavy chain isotype call. The ratio of *IGHA1s*/next highest *IGH* expressed was calculated for each cell across multiple cell lines and transcriptome PCR kits (3′ transcriptome, 5′ GEX and 5′ VDJ). The number of cells that met specific *IGHA1s*/*IGH* ratio minimums were tabulated and presented as a percentage of total population (ratios >2, >3, >4; Table 1). Additionally, the average *IGHA1s*/*IGH* ratio and standard deviation (SD) were calculated for the total population and used to find cells that fell within one or two SD higher than the average ratio; these were presented as a percentage of the total population (1SD, 2SD; Table 1). As shown in Table 1, there were substantial differences between cell lines in the amount of *IGHA1s* expression, but also in the relationship between the SD percentage and the hard ratio cutoff numbers. We did not see a consistent relationship between 1SD or 2SD cell numbers and the hard cutoff ratio numbers (>2, >3, >4), suggesting that biological variability may confound any attempt to assign isotype based on SD.

To assess the overlap of *IGHA1* secreted and membrane splice variants, ratios for *IGHA1s*/*IGHA1m* were calculated for every cell (Figure 7). A pattern similar to that seen in Figure 6 was observed, but with higher ratios associating with higher expression of *IGHA1s*. A significant difference between *IGHA1s*/*m* splice variants and *IGHA1s* and other *IGH* expression is the substantially higher slope value

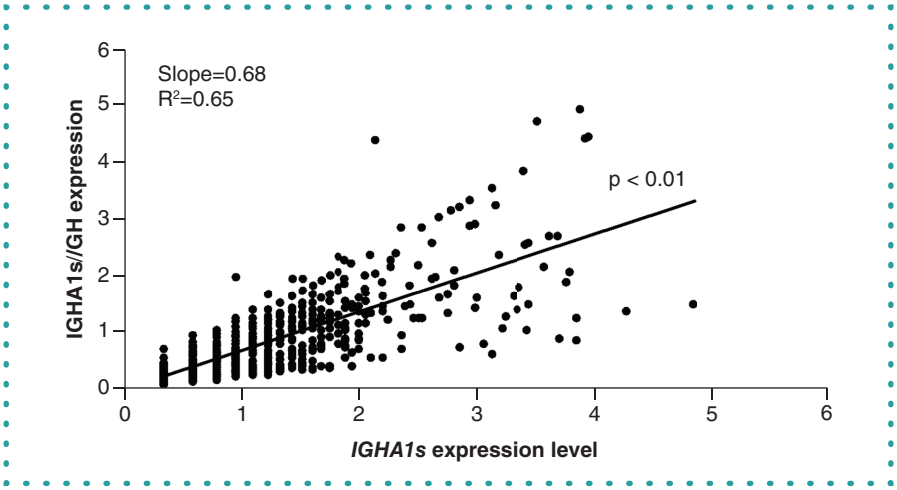

Figure 6. Plot of *IGHA1s*/*IGH* versus *IGHA1s* expression. To determine the relationship of *IGHA1s* expression to other immunoglobulin heavy chains in the same cell, we calculated the slope of *IGHA1s* expression versus the ratio of *IGHA1s*/next highest isotype heavy chain expression in the same cell using the 5′ GEX transcript data. For each cell, all immunoglobulin heavy chain expressions were assessed, and the ratio of *IGHA1s* to the highest or next-highest immunoglobulin heavy chain (*IGHG1*/*2*/*3*/*4, IGHM, IGHD, IGHD*) was calculated. Total number of cells = 2218. Statistical significance was calculated using analysis of variance regression.

Table 1. Percentage of cells out of the total population falling within a specific *IGHA1s*/next-highest isotype heavy chain expression ratio.

5′ GEX	Percentage of total population, *IGHA1s*/*IGH* ratio					*IGHA1s*/*IGH* ratio	
	>2	>3	>4	1SD†	2SD†	Avg	SD
Cell line A	1.76	0.54	0.23	11.45	4.24	0.48	0.49
Cell line B	4.26	1.60	0.70	9.46	4.46	0.63	0.67
Cell line C	6.96	3.19	1.84	8.21	3.77	0.86	0.91
5′ VDJ							
Cell line A	0.45	0.09	0.09	2.93	1.26	0.46	0.48
Cell line B	3.36	1.31	0.66	7.41	3.48	0.62	0.67
Cell line C	0.48	0.14	0.00	1.21	0.48	0.80	0.64
3′ GEX							
Cell line D	8.02	3.66	1.68	9.93	4.50	0.85	0.87
Cell line E	4.06	1.59	0.77	9.72	4.26	0.69	0.63
Cell line F	9.84	4.95	2.72	9.84	4.58	0.93	1.07
Cell line G	3.23	1.44	0.58	9.09	3.69	0.60	0.63

Three cell lines (A, B and C) were analyzed using two different kits: 5′ GEX and 5′ VDJ. Four cell lines (D, E, F and G) were analyzed using the 3′ GEX kit. Each cell was analyzed for its *IGHA1s*/*IGH* ratio using the highest other expressing immunoglobulin heavy chain in that cell. The total number of cells meeting the ratio criteria of >2, >3 and >4 were added up and presented as a percentage of the total population. The average *IGHA1s*/*IGH* ratio for all the cells in the population was calculated, as well as the SD of the total population (right-hand column). These were used to calculate how many cells had *IGHA1s*/*IGH* ratios higher than one or two SD above the average.
†1 or 2 SD higher than the average ratio (Avg + 1*SD or Avg + 2*SD).
Avg: Average; IGH: Immunoglobulin heavy chain isotype; IGHA1s: IgA1-secreting; SD: Standard deviation.

for *IGHA1s*/*IGHA1m* ratios (1.95 vs 0.68; Figures 6 & 7). It is unclear whether this greater specificity in calls for *IGHA1s* versus *IGHA1m* versus delineation from other *IGH* has a technical or biological underpinning. As shown in Table 2, the percentage of cells with high ratios of *IGHA1s*/*IGHA1m* was substantially greater than the comparisons shown in Table 1. This observation further supports the idea that making the call between *IGHA1s* and *IGHA1m* splice variants is easier than for alternative *IGH* expressers when comparing with *IGHA1s*. However, while most antibody-secreting cells do not express cell-surface B cell receptor, there are data suggesting that some antibody-secreting cells have an active B cell receptor [22]. This means that there will be subpopulations of cells that are only IgA1 secretors, some that are only IgA1 presenters and some that are both.

Screening for only cells that have a ratio >2 of *IGHA1s*/*IGH*, we calculated the *IGHA1s*/*IGHA1m* ratio and plotted against *IGHA1s* expression level. This approach provided a mechanism to screen cells for *IGHA1s* specificity versus other *IGH* expressers and determine the splice variants for *IGHA1* secretion, membrane or both within a cell. Figure 8A shows the relationship between *IGHA1s*/*IGHA1m* in cells where the *IGHA1s* expression ratio was >2 compared with expression of all other *IGH* isotypes. The slope (and ratios) in Figure 8A

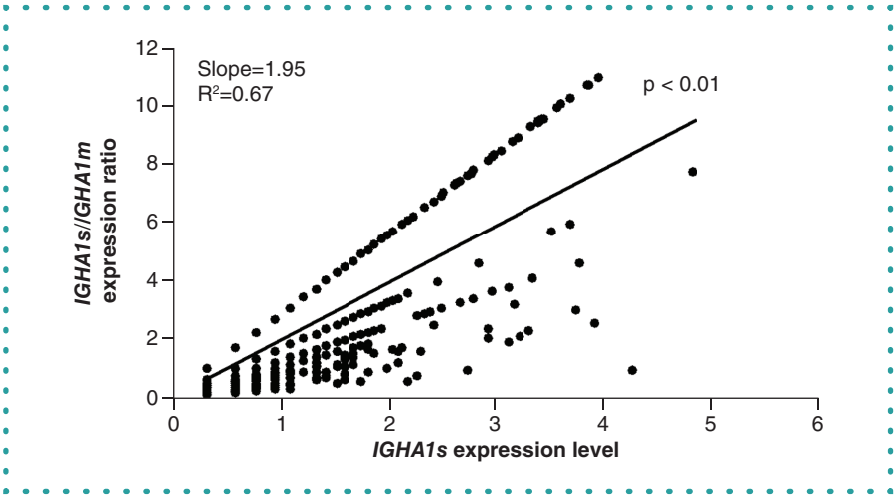

Figure 7. Plot of *IGHA1s*/*IGHA1m* versus *IGHA1s* expression. To determine the relationship of *IGHA1s* expression to *IGHA1m* in the same cell we calculated the slope of *IGHA1s*/*IGHA1m* versus *IGHA1s* expression. *IGHA1s* and *IGHA1m* expression and their ratios in the same cell were calculated from 5′ GEX transcript data. For each cell, *IGHA1s* and *IGHA1m* expressions were assessed, and the ratio of the two was calculated. Total number of cells = 2218. Statistical significance was calculated using analysis of variance regression.

Table 2. Percentage of cells out of the total population falling within a specific *IGHA1s*/*IGHA1m* ratio.

5′ GEX	Percentage of total population, *IGHA1s*/*IGHA1m*					*IGHA1s*/*IGHA1m* ratio	
	>2	>3	>4	1SD†	2SD†	Avg	SD
Cell line A	23.22	11.99	4.82	12.17	4.10	0.48	0.49
Cell line B	16.99	10.81	5.94	9.05	4.50	0.63	0.67
Cell line C	56.62	47.20	20.24	9.95	3.29	0.86	0.91
5′ VDJ							
Cell line A	6.45	3.25	1.35	3.43	1.22	0.46	0.48
Cell line B	13.55	8.56	4.67	8.56	3.81	0.62	0.67
Cell line C	3.96	3.19	1.35	0.72	0.24	0.80	0.64
3′ GEX							
Cell line D	20.09	9.58	5.14	33.71	4.59	0.85	0.87
Cell line E	19.07	7.42	3.92	10.74	4.55	0.69	0.63
Cell line F	16.08	8.23	4.83	9.93	4.83	0.93	1.07
Cell line G	19.94	10.97	6.14	9.92	4.30	0.60	0.63

Three cell lines (A, B and C) were analyzed from two different kits, 5′ GEX and 5′ VDJ. Four cell lines (D, E, F and G) were analyzed from the 3′ GEX kit. Each cell was analyzed for its *IGHA1s*/*IGHA1m* ratio. The total number of cells meeting the ratio criteria of >2, >3 and >4 were added up and presented as a percentage of the total population. The average *IGHA1s*/*IGHA1m* ratio for all the cells in the population was calculated, as well as the SD of the total population (right-hand column). These were used to calculate how many cells had an *IGHA1s*/*IGHA1m* ratio higher than one or two SD above the average.
†1 or 2 SD higher than the average ratio (Avg + 1*SD or Avg + 2*SD).
Avg: Average; IGHA1m: IgA1 membrane-bound; IGHA1s: IgA1-secreting; SD: Standard deviation.

is significantly higher than the slope in the total cell population in *IGHA1s*/*IGHA1m* analysis (Figure 7A), indicating that selecting for lower background of other non-*IGHA1 IGH* transcripts can increase the specificity for *IGHA1s*-expressing cells. Figure 8B shows the relationship of *IGHA1s* to all other *IGH* isotypes expressed in the same cell after screening for a ratio >2. This workflow compares well for isotype specificity (*IGHA1s* in this case) when viewed against the overall isotype overlap in Figure 4C & Figure 6.

An alternate workflow to assess splice variants at the 3′ end using nanopore technology has also been employed with the 10× Genomics 3′ kit. The cDNA was split for gene expression analysis and whole sequencing of antibody genes, providing significant coverage on the 3′ end for secretion versus membrane delineation. This approach requires both nanopore and next-generation sequencing technologies and has a lower recovery of cell barcodes, but could potentially be used alongside a standard 3′ gene expression workflow [7].

Using a combination of R tools to curate the data and Alteryx to visualize and subcategorize cell populations, we can better identify critical IgA1-producing cells with more accurate isotype calls. With this bioinformatic and single-cell transcriptome approach, we can cast a broader net to assess heterogenic populations of IgA1-secreting cells using complex mixtures of immortalized as well as primary B cells.

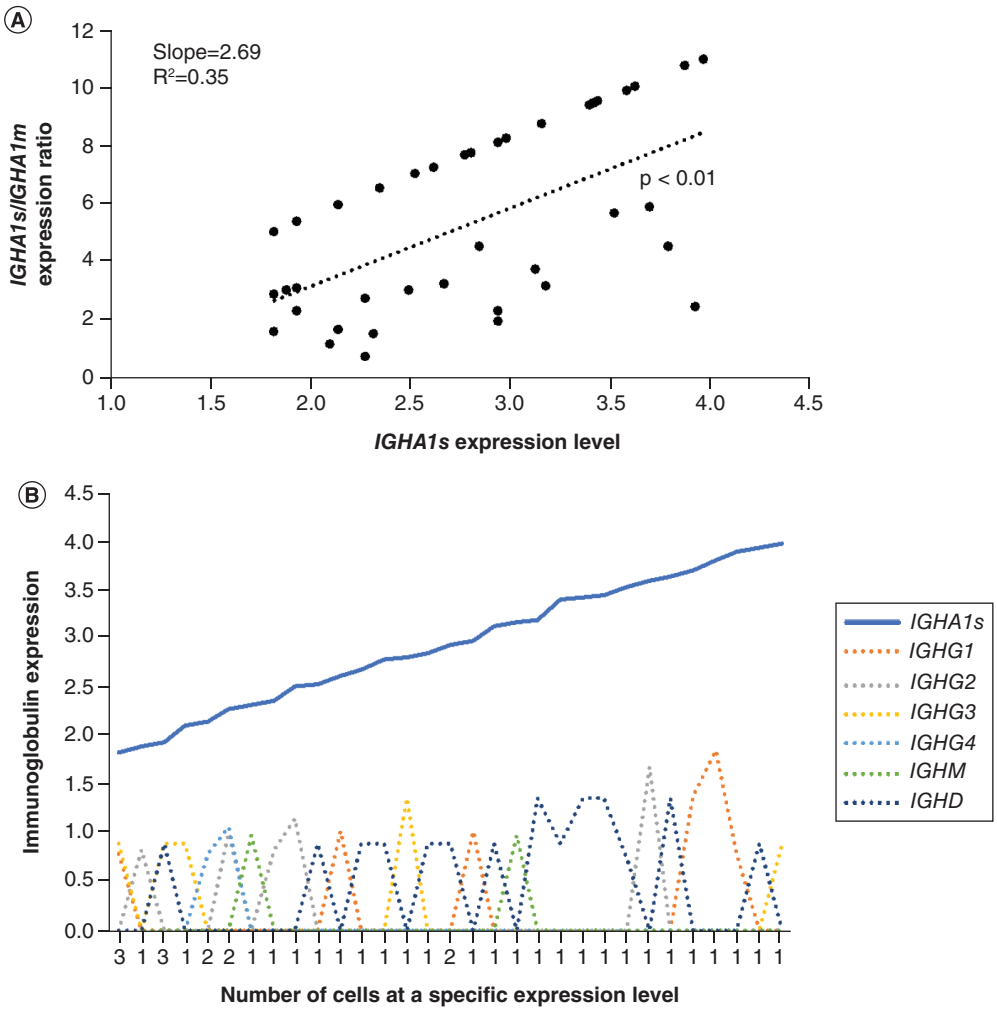

Figure 8. Cells were analyzed for *IGHA1s*/*IGH* chain expression ratios, as depicted in Figure 6 and Table 1, and only cells that had a ratio >2 were used. Those cells were then assessed for *IGHA1s*/*IGHA1m* expression ratios and compared with all other IGH isotypes. **(A)** *IGHA1s* expression level compared with *IGHA1s*/*IGHA1m* ratio for each cell. The x-axis is the *IGHA1s* expression level for each cell and the y-axis is the ratio of *IGHA1s*/*IGHA1m* expression level. **(B)** *IGHA1s* expression level compared with all other IGH isotypes. The X-axis is the number of cells at a specific *IGHA1s* expression level and the y-axis shows *IGHA1s* and IGH isotype expression level. Both panels depict Cell Line A analysis, consisting of 39 out of 2326 cells that met the criteria of >2 ratio versus the next-highest IGH expression level. Statistical significance was calculated using analysis of variance regression.

Future perspective

Future work in primary PBMCs is necessary to assess this bioinformatic workflow in nonimmortalized cells to address potential transcriptional variables associated with the immortalization process. EBV does not infect and immortalize all the B cells in a PBMC pool; thus the immortalized pool only provides a random snapshot of some of the B cells, further highlighting the need for primary cell work. Additionally, to help validate class identification, we will combine immunoglobulin cell-surface detection and single-cell transcriptomics when assessing B cells. One of these technologies used by 10× Genomics is CITE-seq (Cellular Indexing and Transcriptomes and Epitopes), which combines nucleotide-tagged antibodies for cell-surface targets [23]. This work will help to classify some B cells but may be problematic in those antibody-secreting cells that do not present with cell-surface antibodies. This limitation can be addressed by using known subcloned B cells lines that secrete a single isotype.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2020-0044

Author contributions

C Reily designed the study, performed data analysis and drafted the manuscript. N Xu performed data analysis. D Crossman designed the 10× Ranger pipeline and performed data analysis. All authors were responsible for manuscript editing.

Acknowledgments

The authors thank J Novak and B Julian for providing the immortalized cell lines (with IRB approval) and for their advice on this project, and L Shanrun for his support in running and executing the single-cell transcriptomics core at the University of Alabama at Birmingham. C Reily appreciates J Novak and A Agarwal for their mentoring during the K01 award.

Financial & competing interests disclosure

This work was supported by grants from the National Institutes of Health (R03DK122194, K01DK106341) and University of Alabama at Birmingham Center for Clinical and Translational Science (UL1TR003096). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval through the University of Alabama at Birmingham IRB (#070413006). In addition, informed consent was obtained from all donors and participants.

References

1. Wyatt RJ, Julian BA. IgA nephropathy. *N. Engl. J. Med.* 368(25), 2402–2414 (2013).
2. Suzuki H, Moldoveanu Z, Hall S *et al.* IgA1-secreting cell lines from patients with IgA nephropathy produce aberrantly glycosylated IgA1. *J. Clin. Invest.* 118(2), 629–639 (2008).
3. Suzuki H, Raska M, Yamada K *et al.* Cytokines alter IgA1 *O*-glycosylation by dysregulating C1GalT1 and ST6GalNAc-II enzymes. *J. Biol. Chem.* 289, 5330–5339 (2014).
4. Yamada K, Huang ZQ, Raska M *et al.* Inhibition of STAT3 signaling reduces IgA1 autoantigen production in IgA nephropathy. *Kidney Int. Rep.* 2(6), 1194–1207 (2017).
5. Kiryluk K, Li Y, Moldoveanu Z *et al.* GWAS for serum galactose-deficient IgA1 implicates critical genes of the O-glycosylation pathway. *PLoS Genet.* 13(2), e1006609 (2017).
6. Yamada K, Huang Z, Raska M, Reily CR, Aderson J, Suzuki H *et al.* LIF signaling enhances production of galactose-deficient IgA1 in IgA nephropathy. *Kidney Dis.* 6(3), 168–180 (2020).
7. Singh M, Al-Eryani G, Carswell S *et al.* High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* 10(1), 3120 (2019).
8. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36(5), 411–420 (2018).
9. Early P, Hood L. Allelic exclusion and nonproductive immunoglobulin gene rearrangements. *Cell* 24(1), 1–3 (1981).
10. Zheng GX, Lau BT, Schnall-Levin M *et al.* Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* 34(3), 303–311 (2016).
11. Lindeman I, Emerton G, Mamanova L *et al.* BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods* 15(8), 563–565 (2018).
12. See P, Lum J, Chen J, Ginhoux F. A single-cell sequencing guide for immunologists. *Front. Immunol.* 9, 2425 (2018).
13. Alteryx, Inc. http://www.alteryx.com
14. Satija Lab. Seurat software. https://satijalab.org/seurat/
15. RStudio. https://rstudio.com
16. Sinclair AJ, Farrell PJ. Host cell requirements for efficient infection of quiescent primary B lymphocytes by Epstein–Barr virus. *J. Virol.* 69(9), 5461–5468 (1995).
17. Owen JA, Punt J, Stranford SA, Jones PP, Kuby J. *Kuby Immunology (7th Edition)*. WH Freeman, NY, USA (2013).
18. Shi Z, Zhang Q, Yan H *et al.* More than one antibody of individual B cells revealed by single-cell immune profiling. *Cell Discov.* 5, 64 (2019).
19. 10× Genomics. What is Loupe browser? https://support.10×genomics.com/single-cell-gene-expression/software/visualization/latest/what-is-loupe-cell-browser
20. Janeway CA, Travers P, Walport M, Schlomchik MJ. *Structural Variation in Immunoglobulin Constant Regions*. Garland Science, NY, USA (2001).
21. Hu Q, Hong Y, Qi P *et al.* An atlas of infiltrated B-lymphocytes in breast cancer revealed by paired single-cell RNA-sequencing and antigen receptor profiling. *bioRxiv* doi:10.1101/695601 (2019). [Epub ahead of print]
22. Pinto D, Montani E, Bolli M *et al.* A functional BCR in human IgA and IgM plasma cells. *Blood* 121(20), 4110–4114 (2013).
23. Stoeckius M, Hafemeister C, Stephenson W *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14(9), 865–868 (2017).