An RNA-seq primer for pulmonologists

Sarah G. Chu¹, Sergio Poli De Frias ¹, Benjamin A. Raby^{1,2,3} and Ivan O. Rosas¹

Affiliations: ¹Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ²Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ³Division of Pulmonary and Respiratory Diseases, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.

Correspondence: Ivan O. Rosas, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA. E-mail: irosas@rics.bwh.harvard.edu

@ERSpublications

RNA sequencing is a powerful tool for high-resolution transcriptomic analysis and can be leveraged to better understand the molecular underpinnings of diverse lung diseases <http://bit.ly/2mqBnl4>

Cite this article as: Chu SG, Poli De Frias S, Raby BA, *et al.* An RNA-seq primer for pulmonologists. *Eur Respir J* 2020; 55: 1801625 [<https://doi.org/10.1183/13993003.01625-2018>].

Introduction

With the evolution of high throughput sequencing technologies, the past decade has seen an exponential rise in the use of RNA sequencing (RNA-seq). RNA-seq has deepened our understanding of biological systems to unprecedented levels of resolution, identifying not only gene expression signatures but also regulatory RNA molecules that may play critical roles in disease pathogenesis. Pulmonary research has quickly incorporated this technology, from characterising the IL-17 signature of steroid-unresponsive COPD patients [1] to discovering pathogen-host interactions of *Mycobacterium tuberculosis* [2]. This mini-review provides an overview of core features and applications of RNA-seq to familiarise non-experts with the methodology and how it has impacted our understanding of lung pathophysiology.

Basic principles of RNA-seq

Messenger and non-coding RNA together comprise the transcriptome, a dynamic representation of the functional elements of a cell or organism's genome in a given physiological condition [3]. Earlier methods to analyse the transcriptome consisted of sequencing short segments of complementary DNA (cDNA) with Sanger-based methods or using gene expression microarrays, in which cDNA is quantified by hybridisation with known probes. These approaches were limited by time, incomplete characterisation of sequences, background noise, and reliance on reference genomes [3]. In contrast, RNA-seq applies next generation sequencing to read tens to hundreds of millions of transcripts simultaneously, capturing subtle genetic variations and known or novel RNA isoforms with or without a reference genome.

While several RNA-seq platforms are available, most share a common pathway consisting of isolating RNA, generating cDNA libraries, amplifying and sequencing cDNA, and analysing data (figure 1).

Experimental design

Biological, technical and statistical biases can be minimised by thoughtful experimental design with clearly defined objectives and performing quality control assessments at each stage of the RNA-seq pipeline [4]. Selection of the specimen to be sequenced is guided by the question being asked; induced sputum [5], peripheral blood cells [6] and bronchoalveolar lavage fluid [7] may be useful for surveying the immune landscape and identifying potential biomarkers less invasively, for example yielding important insights into

Received: 26 Aug 2018 | Accepted after revision: 16 Sept 2019

Copyright ©ERS 2020

FIGURE 1 RNA-seq pipeline: A typical RNA sequencing experiment consists of designing the experiment, isolating RNA from the desired cells or tissue, generating cDNA libraries prior or following fragmentation, sequencing cDNA, and processing and analysing data.

the Th2 signatures of asthmatics [6], whereas structural mechanisms may be better interrogated with nasal or bronchial brushings [8, 9], biopsies [10] or lung explants [11, 12]. Nasal brushings are being increasingly used as they are more accessible than bronchial specimens, provide high quality RNA [9] and may be used as a surrogate for bronchial epithelial cells in functional studies [13]. Isolating specific cell types from heterogeneous tissue can be achieved by flow-based strategies [14, 15] but are often constrained by marker specificity. Cultured primary cells may not retain the characteristics of parent tissue and animal- and cell line-based models may provide only limited insights into human disease, yet these may be more accessible and controlled than human samples. Each approach also poses unique technical challenges; sputum collection times can influence microbiotic diversity [16] while explanted tissue is often subject to long ischaemic times, potentially affecting RNA yield and quality [17]. RNA preservation medium can improve yield and quality, and novel techniques have been developed to increase biomarker concentration from bronchoscopy specimens [18]. Table 1 highlights some of the key features of the main respiratory specimen types.

Another consideration is whether the information sought is primarily quantitative (e.g. assessing differential gene expression in lung endothelial cells in idiopathic pulmonary arterial hypertension (PAH) [11]) or qualitative (e.g. identifying fusion genes in lung cancers [19]). This determines the need for biological replicates and appropriate sequencing coverage (portion of targeted region being sequenced) and depth (number of independent reads) [20]. Read depth is a critical determinant of sequencing accuracy and the ability to identify sequence variants. Appropriate coverage and depth may vary widely depending on the cell type or developmental stage [21]. Long reads, paired-end reads, and strand specificity are important for *de novo* transcript assembly and isoform classification but come at a trade-off of increased cost and reduced depth [22, 23].

RNA isolation, library preparation and sequencing

Obtaining adequate RNA while preserving integrity is critical for successful RNA-seq [20]. This can be challenging, as RNA is notoriously susceptible to degradation especially in clinical or formalin-fixed paraffin-embedded specimens. The advantages of fresh samples must be weighed against biases introduced by batch effects. A commonly used metric of quality is the RNA Integrity Number (RIN), which reflects RNA degradation as a ratio of 28S and 18S rRNA but is less informative regarding sample purity. RIN scores of 7 or greater are often recommended for RNA-seq. Ultraviolet absorbance is a rapid method of assessing RNA quality and concentration but lacks specificity for RNA and does not directly assess degradation.

Extraction is carried out with silica gel membrane columns or organic solvents (which must be subsequently removed to prevent interference with downstream steps, including reverse transcription). Transcript RNA can be enriched from total RNA by poly(A) selection using oligo(dT) primers and depletion of ribosomal RNA (rRNA) with enzymatic digestion or probe-based elimination. A high

TABLE 1 Main specimen types that have been used for respiratory RNA sequencing studies and technical considerations

Specimen type	Disease	Technical considerations
Sputum	Asthma [5], CF [48]	Immune cells; contamination by bacterial DNA, difficult to homogenise; RIN 7.4–7.7
Peripheral blood	Asthma [6], TB [49], sarcoidosis [50], COPD [51]	Immune cells; high RIN [52], easy to obtain, lower specificity for lung diseases
Nasal brushing	Asthma [9]	Epithelial cells; high RIN but may vary depending on location (inferior turbinate <i>versus</i> anterior nare) [9, 53]
Bronchial lavage fluid	Lung allograft dysfunction [7], COPD [51]	Immune cells; variable RIN (5.2–7.7) [7]
Bronchial brushing	Smokers [8], COPD [51, 54]	Epithelial cells; RIN not dependent on airway location or subject phenotype [55]
Transbronchial biopsies, needle aspiration	Asthma [10], IPF [46], lung cancer [56]	Gene expression associated with RNA quality and quantity [57]
Lung tissue	IPF [12, 15, 44], COPD [12, 58]	Heterogeneous cell composition, wide variability of RIN, not significantly affected by cryopreservation

CF: cystic fibrosis; TB: tuberculosis; IPF: idiopathic pulmonary fibrosis; RIN: RNA integrity number.

proportion of poly(A) tails may be lost in degraded RNA; to avoid underrepresenting the transcriptome, alternative enrichment approaches should be considered, such as exome capture [17, 24]. Enriched RNA is fragmented to fit the size range of the sequencing platform.

Reverse transcription generates cDNA inserts which are then ligated to adapters; these serve as amplification signals and sample-specific barcodes. cDNA libraries are usually amplified by PCR before sequencing. Commonly used commercial sequencing platforms include Illumina, SOLiD, PacBio and Ion Torrent, which vary based on library preparation, sequencing chemistry and the characteristics of data generated, such as maximum sequence read length (*e.g.* PacBio can produce significantly longer reads). To minimise cost and time, high throughput sequencers can combine (*i.e.* “multiplex”) libraries from several experiments into a single sequencing reaction.

Single-cell RNA-seq

Single-cell RNA-seq (scRNA-seq) has rapidly gained popularity to assess gene expression in individual cells. Whereas bulk RNA-seq profiles transcriptomes averaged across a pool of cells, scRNA-seq can resolve mixed populations including those in tumours, complex organs, and progenitor and immune cells. For example, scRNA-seq delineated 52 stromal cell subtypes in the lung tumour microenvironment [25] and identified aberrant alveolar epithelial subtypes associated with idiopathic pulmonary fibrosis (IPF) [15]. Recently, scRNA-seq and lineage tracing of human and mouse airway cells revealed “pulmonary ionocytes”, a novel epithelial cell critical for CFTR activity [26, 27].

ScRNA-seq employs a powerful amplification process to sequence transcripts from minute quantities of RNA. Advanced microfluidics (*e.g.* Drop-seq) enable separation of tens of thousands of cells into nanolitre-sized droplets [28]. Nuc-seq has further developed this technology to sequence single-nuclear RNA from frozen tissue [29]. Clustering similar cells is important for generating a more comprehensive transcriptome and reducing dimensionality [30]. Major challenges of scRNA-Seq include cost, low gene expression levels and ensuring complete dissociation of cells, particularly in solid tissues. This limits sequencing depth and the capacity to detect rare transcripts [20, 30].

Data processing and analysis

Processing the “big data deluge” inherent to omic studies can be daunting, requiring thoughtful selection from a broad range of bioinformatic tools. After obtaining raw reads, sequencing artefacts (*e.g.* from adapters and very short or low complexity reads) are filtered out. The next step, aligning reads to a reference genome or transcriptome [31], is perhaps the most computationally demanding; each read must be accurately and uniquely mapped, considering sequencing errors *versus* true biological variation (*i.e.* insertions, deletions or single nucleotide polymorphisms) [32]. Sophisticated algorithms can reconstruct full transcripts, assemble *de novo* transcripts, quantitate gene expression, examine small RNA species and long non-coding RNAs, and detect alternative splicing, gene fusion and variants [33, 34]. Commonly used RNA-seq aligners include TopHat2, STAR and HISAT. Alignment accuracy can vary widely depending on factors including the number of intron-spanning reads and transcript annotation [35]. The percentage of mapped reads is an important metric of overall sequencing quality. While unmapped reads often reflect

sequencing errors or contamination by rRNA or genomic DNA, they may also represent unannotated transcripts or true biological variants [36]. Conversely, multimapped reads can result from isoforms of identical sequences [20]. Reads are quantified and normalised by transcript length and library size.

With normalised quantitative data in hand, one can now ask: what are the differences in gene expression between states? Which pathways are activated? These questions can be answered in a multitiered way usually beginning with differential gene expression. Genes are ranked by the strength of the signal and p-values adjusted for multiple testing using statistical packages such as DESeq, edgeR or limma/voom, which differ based on the distribution model and treatment of sample size, replicates, sequencing depth and overdispersion (*i.e.* observed *versus* theoretical dispersion). For example, DESeq2 models data using negative binomial distribution and Bayes theorem to estimate dispersion [37], whereas limma/voom employs linear modelling and estimates the mean–variance relationship of gene counts to weigh each observation [38]. Comparisons of these software have been studied systematically [39, 40].

Because individual genes may exert only a weak phenotypic effect, pathway enrichment analysis with databases of functionally annotated gene sets (*e.g.* GO, KEGG or Reactome) can be used to group differentially expressed genes based on biological processes or molecular functions. This not only highlights key regulatory pathways but reduces the dimensionality of the statistical analysis [41]. Co-expression network analysis, which identifies networks of highly correlated genes to infer gene function and gene-disease associations [42], and integration with other omic datasets including genomic, methylation, microRNA, proteomic and metabolomic studies, can complement RNA-seq to model biologically correlated behaviours [20, 43]. For example, parallel microRNA and mRNA-seq data from IPF and COPD lungs were integrated to create a gene/miRNA regulatory network, identifying MIR96 as a regulator of the p53/hypoxia lung response to environmental injury [12].

Experimental validation

RNA-seq is a powerful tool for discovery, but must be augmented by other experimental approaches to validate findings, establish causality and elucidate mechanism. Gene signatures may be validated at the protein expression level and further investigated with cell- and animal-based functional assays. After combining RNA-seq and microarray findings to identify senescence markers in IPF lung, SCHAFER *et al.* [44] assessed p16 expression in fibrotic lung tissue *via* immunostaining, characterised the profibrotic secretome in senescent fibroblasts, and then found that clearance of senescent cells improves lung compliance in experimental fibrosis. EFNA1, a gene related to BMPR2 signalling found to be downregulated in human PAH endothelial cells by RNA-seq, was knocked down *in vitro* and in mice to recapitulate PAH phenotypes [11]. Finally, as an example of how RNA-seq targets might be translated directly to clinical therapies, the discovery of PD-1-positive progenitor cells of type 2 innate lymphoid cells (ILC2s) by scRNA-seq led to reduction of acute lung inflammation in mice treated with antibody-depleted PD-1^{hi} ILCs [45] and highlighted the potential for PD-1 inhibitors, already widely used in cancers, to treat ILC2-mediated disorders such as allergic asthma.

Such studies underscore the capacity of RNA-seq to generate hypotheses and uncover novel therapeutic targets. It has thus emerged as the method of choice for transcriptome analysis, although challenges, including cost (typically several hundred dollars per sample, and higher for scRNA-seq, greater sequencing depths and lengths and more sensitive platforms) remain. Moreover, the volume and complexity of data requires a level of bioinformatic and statistical expertise currently in short supply. Despite its enormous potential, direct clinical applications of RNA-seq remain limited. RNA-seq has been integrated with machine learning to diagnose IPF from transbronchial biopsies [46] and mild/moderate asthma from nasal brushings [9]. Other RNA-based technologies, primarily using quantitative real-time PCR and microarrays, are already employed in the clinical setting, including the detection of respiratory RNA viral agents (*e.g.* severe acute respiratory syndrome and Middle East respiratory syndrome) and for lung cancer diagnosis and prognostication [47]. With the strong foothold RNA-seq provides for resolving complex gene structures and functions, future efforts should be directed not only towards gaining a comprehensive understanding of lung pathobiology, but also translating discoveries to clinically useful biomarkers and targets for precision therapies.

Conflict of interest: S.G. Chu has nothing to disclose. S. Poli De Frias has nothing to disclose. B.A. Raby reports grants from National Institutes of Health (P01HL114501), during the conduct of the study. I.O. Rosas reports grants from National Institutes of Health (P01HL114501), during the conduct of the study.

References

- 1 Christenson SA, van den Berge M, Faiz A, *et al.* An airway epithelial IL-17A response signature identifies a steroid-unresponsive COPD patient subgroup. *J Clin Invest* 2019; 129: 169–181.
- 2 Rienksma RA, Suarez-Diez M, Mollenkopf HJ, *et al.* Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. *BMC Genomics* 2015; 16: 34.
- 3 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; 10: 57–63.

- 4 Consortium SM-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* 2014; 32: 903–914.
- 5 Jones AC, Troy NM, White E, *et al.* Persistent activation of interlinked type 2 airway epithelial gene networks in sputum-derived cells from aeroallergen-sensitized symptomatic asthmatics. *Sci Rep* 2018; 8: 1511.
- 6 Seumois G, Zapardiel-Gonzalo J, White B, *et al.* Transcriptional profiling of Th2 cells identifies pathogenic features associated with asthma. *J Immunol* 2016; 197: 655–664.
- 7 Weigt SS, Wang X, Palchevskiy V, *et al.* Gene expression profiling of bronchoalveolar lavage cells preceding a clinical diagnosis of chronic lung allograft dysfunction. *PLoS One* 2017; 12: e0169894.
- 8 Hackett NR, Butler MW, Shaykhiev R, *et al.* RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics* 2012; 13: 82.
- 9 Pandey G, Pandey OP, Rogers AJ, *et al.* A nasal brush-based classifier of asthma identified by machine learning analysis of nasal RNA sequence data. *Sci Rep* 2018; 8: 8826.
- 10 Yick CY, Zwinderman AH, Kunst PW, *et al.* Transcriptome sequencing (RNA-Seq) of human endobronchial biopsies: asthma *versus* controls. *Eur Respir J* 2013; 42: 662–670.
- 11 Rhodes CJ, Im H, Cao A, *et al.* RNA sequencing analysis detection of a novel pathway of endothelial dysfunction in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2015; 192: 356–366.
- 12 Kusko RL, Brothers JF 2nd, Tedrow J, *et al.* Integrated genomics reveals convergent transcriptomic networks underlying chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 194: 948–960.
- 13 Brewington JJ, Filbrandt ET, LaRosa FJ 3rd, *et al.* Brushed nasal epithelial cells are a surrogate for bronchial epithelial CFTR studies. *JCI Insight* 2018; 3: 99385.
- 14 Fujino N, Kubo H, Ota C, *et al.* A novel method for isolating individual cellular components from the adult human distal lung. *Am J Respir Cell Mol Biol* 2012; 46: 422–430.
- 15 Xu Y, Mizuno T, Sridharan A, *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: e90558.
- 16 Cuthbertson L, Rogers GB, Walker AW, *et al.* Time between collection and storage significantly influences bacterial sequence composition in sputum samples from cystic fibrosis respiratory infections. *J Clin Microbiol* 2014; 52: 3011–3016.
- 17 Chu SG, Poli De Frias S, Sakairi Y, *et al.* Biobanking and cryopreservation of human lung explants for omic analysis. *Eur Respir J* 2020; 55: 1801635.
- 18 Leaker BR, Nicholson GC, Ali FY, *et al.* Bronchoabsorption; a novel bronchoscopic technique to improve biomarker sampling of the airway. *Respir Res* 2015; 16: 102.
- 19 Fernandez-Cuesta L, Sun R, Menon R, *et al.* Identification of novel fusion genes in lung cancer using breakpoint assembly of transcriptome sequencing data. *Genome Biol* 2015; 16: 7.
- 20 Conesa A, Madrigal P, Tarazona S, *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016; 17: 13.
- 21 Sims D, Sudbery I, Illott NE, *et al.* Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014; 15: 121–132.
- 22 McIntyre LM, Lopiano KK, Morse AM, *et al.* RNA-seq: technical variability and sampling. *BMC Genomics* 2011; 12: 293.
- 23 Levin JZ, Yassour M, Adiconis X, *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* 2010; 7: 709–715.
- 24 Cieslik M, Chugh R, Wu YM, *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res* 2015; 25: 1372–1381.
- 25 Lambrechts D, Wauters E, Boeckx B, *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* 2018; 24: 1277–1289.
- 26 Montoro DT, Haber AL, Biton M, *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018; 560: 319–324.
- 27 Plasschaert LW, Zilionis R, Choo-Wing R, *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018; 560: 377–381.
- 28 Macosko EZ, Basu A, Satija R, *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015; 161: 1202–1214.
- 29 Krishnaswami SR, Grindberg RV, Novotny M, *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* 2016; 11: 499–524.
- 30 Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol* 2018; 14: 479–492.
- 31 Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* 2012; 22: 271–274.
- 32 Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. *Genome Biol* 2010; 11: 220.
- 33 Trapnell C, Williams BA, Pertea G, *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; 28: 511–515.
- 34 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25: 1105–1111.
- 35 Baruzzo G, Hayer KE, Kim EJ, *et al.* Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods* 2017; 14: 135–139.
- 36 van der Weide RH, Simonis M, Hermesen R, *et al.* The genomic scrapheap challenge; extracting relevant data from unmapped whole genome sequencing reads, including strain specific genomic segments, in rats. *PLoS ONE* 2016; 11: e0160036.
- 37 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15: 550.
- 38 Ritchie ME, Phipson B, Wu D, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
- 39 Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* 2017; 12: e0190152.

- 40 Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinformatics* 2015; 16: 59–70.
- 41 Clark NR, Szymkiewicz M, Wang Z, *et al.* Principal angle enrichment analysis (PAEA): dimensionally reduced multivariate gene set enrichment analysis tool. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2015; 2015: 256–262.
- 42 van Dam S, Vosa U, van der Graaf A, *et al.* Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinformatics* 2018; 19: 575–592.
- 43 Hung JH. Gene set/pathway enrichment analysis. *Methods Mol Biol* 2013; 939: 201–213.
- 44 Schafer MJ, White TA, Iijima K, *et al.* Cellular senescence mediates fibrotic pulmonary disease. *Nat Commun* 2017; 8: 14532.
- 45 Yu Y, Tsang JC, Wang C, *et al.* Single-cell RNA-seq identifies a PD-1(hi) ILC progenitor and defines its development pathway. *Nature* 2016; 539: 102–106.
- 46 Choi Y, Lu J, Hu Z, *et al.* Analytical performance of Envisia: a genomic classifier for usual interstitial pneumonia. *BMC Pulm Med* 2017; 17: 141.
- 47 Byron SA, Van Keuren-Jensen KR, Engelthaler DM, *et al.* Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* 2016; 17: 257–271.
- 48 Rossi E, Falcone M, Molin S, *et al.* High-resolution *in situ* transcriptomics of *Pseudomonas aeruginosa* unveils genotype independent patho-phenotypes in cystic fibrosis lungs. *Nat Commun* 2018; 9: 3459.
- 49 Thompson EG, Du Y, Malherbe ST, *et al.* Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis (Edinb)* 2017; 107: 48–58.
- 50 Talreja J, Farshi P, Alazizi A, *et al.* RNA-sequencing identifies novel pathways in sarcoidosis monocytes. *Sci Rep* 2017; 7: 2720.
- 51 Morrow JD, Chase RP, Parker MM, *et al.* RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD. *Respir Res* 2019; 20: 65.
- 52 Kozlakidis Z, Mant C, Abdinur F, *et al.* Variation of peripheral blood mononuclear cell RNA quality in archived samples. *Biopreserv Biobank* 2011; 9: 259–263.
- 53 Lai PS, Liang L, Cibas ES, *et al.* Alternate methods of nasal epithelial cell sampling for airway genomic studies. *J Allergy Clin Immunol* 2015; 136: 1120–1123 e1124.
- 54 Yeo J, Morales DA, Chen T, *et al.* RNAseq analysis of bronchial epithelial cells to identify COPD-associated genes and SNPs. *BMC Pulm Med* 2018; 18: 42.
- 55 Raman T, O'Connor TP, Hackett NR, *et al.* Quality control in microarray assessment of gene expression in human airway epithelium. *BMC Genomics* 2009; 10: 493.
- 56 Kage H, Kohsaka S, Shinozaki-Ushiku A, *et al.* Small lung tumor biopsy samples are feasible for high quality targeted next generation sequencing. *Cancer Sci* 2019; 110: 2652–2657.
- 57 Pankratz DG, Choi Y, Imtiaz U, *et al.* Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. *Ann Am Thorac Soc* 2017; 14: 1646–1654.
- 58 Kim WJ, Lim JH, Lee JS, *et al.* Comprehensive analysis of transcriptome sequencing data in the lung tissues of COPD subjects. *Int J Genomics* 2015; 2015: 206937.