


RESEARCH ARTICLE

Open Access



Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis

Milica Vukmirovic^{7*†} , Jose D. Herazo-Maya^{7†}, John Blackmon¹, Vesna Skodric-Trifunovic^{2,3}, Dragana Jovanovic^{2,3}, Sonja Pavlovic⁴, Jelena Stojic⁵, Vesna Zeljkovic⁶, Xiting Yan⁷, Robert Homer^{8,9}, Branko Stefanovic^{1†} and Naftali Kaminski^{7†}

Abstract

Background: Idiopathic Pulmonary Fibrosis (IPF) is a lethal lung disease of unknown etiology. A major limitation in transcriptomic profiling of lung tissue in IPF has been a dependence on snap-frozen fresh tissues (FF). In this project we sought to determine whether genome scale transcript profiling using RNA Sequencing (RNA-Seq) could be applied to archived Formalin-Fixed Paraffin-Embedded (FFPE) IPF tissues.

Results: We isolated total RNA from 7 IPF and 5 control FFPE lung tissues and performed 50 base pair paired-end sequencing on Illumina 2000 HiSeq. TopHat2 was used to map sequencing reads to the human genome. On average ~62 million reads (53.4% of ~116 million reads) were mapped per sample. 4,131 genes were differentially expressed between IPF and controls (1,920 increased and 2,211 decreased (FDR < 0.05). We compared our results to differentially expressed genes calculated from a previously published dataset generated from FF tissues analyzed on Agilent microarrays (GSE47460). The overlap of differentially expressed genes was very high (760 increased and 1,413 decreased, FDR < 0.05). Only 92 differentially expressed genes changed in opposite directions. Pathway enrichment analysis performed using MetaCore confirmed numerous IPF relevant genes and pathways including extracellular remodeling, TGF-beta, and WNT. Gene network analysis of MMP7, a highly differentially expressed gene in both datasets, revealed the same canonical pathways and gene network candidates in RNA-Seq and microarray data. For validation by NanoString nCounter[®] we selected 35 genes that had a fold change of 2 in at least one dataset (10 discordant, 10 significantly differentially expressed in one dataset only and 15 concordant genes). High concordance of fold change and FDR was observed for each type of the samples (FF vs FFPE) with both microarrays ($r = 0.92$) and RNA-Seq ($r = 0.90$) and the number of discordant genes was reduced to four.

Conclusions: Our results demonstrate that RNA sequencing of RNA obtained from archived FFPE lung tissues is feasible. The results obtained from FFPE tissue are highly comparable to FF tissues. The ability to perform RNA-Seq on archived FFPE IPF tissues should greatly enhance the availability of tissue biopsies for research in IPF.

Keywords: Idiopathic Pulmonary Fibrosis, FFPE, RNA-Seq, Microarray, DEGs, Validation, Pathways, Network, MMP7, NanoString nCounter[®]

* Correspondence: milica.vukmirovic@yale.edu

†Equal contributors

⁷Section of Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine, New Haven, CT, USA

Full list of author information is available at the end of the article



Background

Idiopathic Pulmonary Fibrosis (IPF) is a chronic interstitial lung disease of unknown etiology associated with high mortality rates, and increased prevalence with age (14–42.7 cases per 100,000 population) [1]. IPF patients have an overall median survival of approximately 3.5 years from the onset of symptoms. The disease is characterized by progressive scarring of the lung parenchyma that leads to loss of lung function [2]. IPF is thought to result from repeated cycles of alveolar epithelial cell injury leading to fibroblast proliferation, exaggerated accumulation of extracellular matrix in the lung parenchyma and recapitulation of developmental pathways [3, 4].

Currently, IPF diagnosis is based on the exclusion of known causes of lung fibrosis, and the presence of a Usual Interstitial Pneumonia (UIP) pattern on High-Resolution Computed Tomography scan (HRCT) in patients who do not undergo lung biopsy or the combination of a permissive HRCT and the UIP pattern on surgical lung biopsy [5]. The UIP histology pattern is characterized by spatial and temporal heterogeneity, which refers to patchy distribution of dense parenchymal scar along with areas of fibroblast and myofibroblast accumulation and proliferation, known as fibroblastic foci, alternating with areas of less affected or normal lung parenchyma [5, 6]. While lung biopsies were frequently performed to identify the typical UIP pattern as part of the IPF diagnostic workup, the success of HRCT in demonstrating “UIP” radiological patterns has considerably limited the number of lung biopsies currently being performed [7]. That in turn has led to a decline in the availability of tissues for IPF research in general, and for transcriptomic profiling in particular.

Transcriptomic profiling in IPF has largely been performed by microarrays using RNA obtained from whole lung lysates obtained from fresh frozen tissues. These studies have provided significant mechanistic insights regarding IPF pathogenesis, and have largely impacted the field of lung fibrosis [4, 8]. However, these studies are limited because acquisition of fresh frozen tissues is only available in highly specialized academic centers with tissue banking facilities. Thus, the majority of studies contain mostly tissues explanted from patients with IPF at the time of biopsy, and studies containing tissues obtained from diagnostic biopsies are limited. Additionally, it is nearly impossible to assess the lung morphology on frozen tissue, thus the studies utilizing fresh frozen samples depend on histological assessment of adjacent tissue which may or may not contain the exact same pathology. While transcriptomic data generated from different dissections within a single lobe of the lung are highly correlated [9], and that transcriptomic data correlates well with UIP pattern itself [10, 11] the lack of visual confirmation of the histology of the region profiled is still considered a limitation [12].

RNA isolated from Formalin-Fixed Paraffin Embedded FFPE tissues is partially degraded, thus transcriptomic analysis of FFPE tissues was considered challenging [13, 14]. Several recent studies demonstrated that transcriptomic analysis of FFPE tissues using microarrays was possible, was nearly comparable to fresh frozen tissues but still had significant limitations [15–17]. In contrast to microarrays, next generation RNA Sequencing (RNA-Seq) allows for relatively unbiased measurements of expression levels across the entire length of a transcript and its level of expression [18], and therefore may be more suitable for sequencing of partially degraded FFPE RNA. Transcriptomic analysis of FFPE tissues by RNA-Seq demonstrates high concordance to RNA-Seq data produced from matching fresh frozen tissues [19–23]. Because formalin fixation and paraffin embedding is routinely done on all samples from clinically indicated lung biopsies, optimization of a method to perform genome scale transcript profiling of archived FFPE tissues will greatly enhance the access to IPF lungs.

In this study, we sought to determine whether whole transcriptomic analysis of RNA isolated from FFPE biopsies by RNA-Seq was feasible in IPF, and whether the results are comparable to those obtained from gene expression microarrays. To test our hypothesis, we isolated RNA from FFPE lung biopsies of IPF individuals and controls, generated RNA-Seq expression data and compared it to publically available microarray data previously generated by us from fresh frozen IPF lung tissues (GSE47460, [24]) (Fig. 1). Our study demonstrates high concordance in IPF relevant genes and pathways between RNA-Seq and microarrays of un-paired tissues from patients evaluated in different cohorts, suggesting that RNA-Seq from FFPE tissues could be considered an acceptable technique for transcriptomic profiling in IPF.

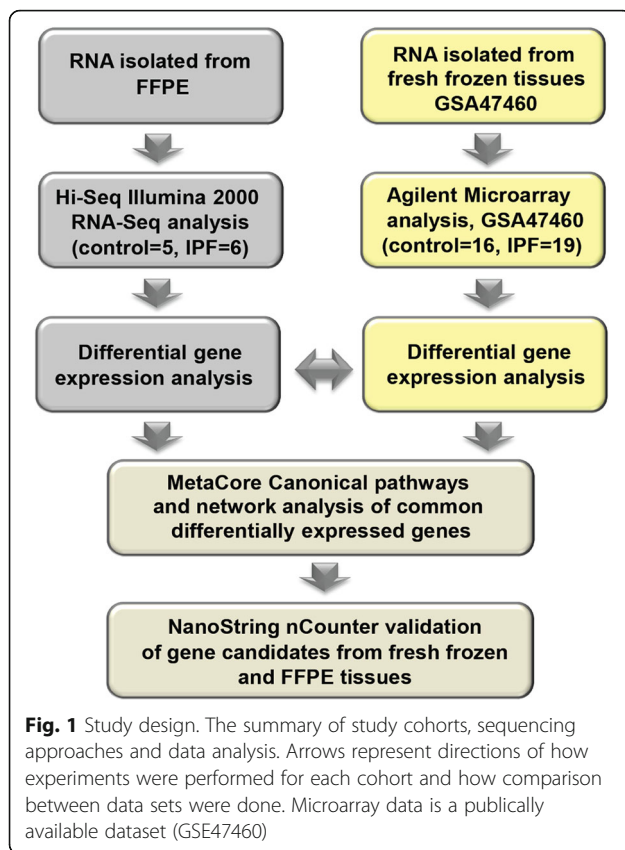
Methods

FFPE Tissue specimens

Lung FFPE biopsies were obtained from departmental FFPE archives [Clinic for Pulmonology in Belgrade, $n = 8$, and the Lung Tissue Research Consortium (LTRC), $n = 4$] according to Institutional Review Board (IRB) approved study protocols. Informed consents to participate in the study were also obtained according to IRB. The median archival age of FFPE tissues was 6 years. IPF ($n = 7$) lung and control ($n = 5$) FFPE tissues were used for RNA-Seq. Clinical, demographic and histopathological features of the subjects in the study were analyzed by a multidisciplinary group of clinicians and pathologists to confirm IPF diagnosis [5].

FFPE RNA isolation and quality control

Five 10- μ m slices of the whole lung tissue were cut from FFPE block, excess paraffin was trimmed and slices were treated twice with 1ml xylene for 30 min at 62°C, then



washed twice with 100% ethanol as previously described [25]. Total RNA was isolated by using MasterPure kit (Epicentre Biotechnologies). The final RNA concentration and purity (A_{260}/A_{280}) was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). RNA quality and RNA integrity (RIN) was assessed using a 2100 Bioanalyzer (Agilent). For each FFPE tissue block, two consecutive RNA isolations from the whole lung tissue slices ($5 \times 10\text{-}\mu\text{m}$) were performed.

Fresh frozen tissue specimens

The Lung Genomic Research Consortium (LGRC) contains data for over 200 IPF patients and controls. To have similar size cohorts we picked 35 age and gender matching samples (19 IPF and 16 normal histology samples) and reanalyzed the data. The experiments were approved by the Institutional Review Board. These samples are publically available at GSE47460 and at the LGRC website - <http://www.lung-genomics.org>) and have been previously described at [24, 26].

RNA-Seq library preparation and paired-end RNA-Seq

Approximately $1.5\mu\text{g}$ of total RNA was isolated from each FFPE block (tissues size $\sim 10\text{mm} \times 7\text{mm}$). FFPE RNA showed fragmentation in a range of $\sim 100\text{-}150\text{bp}$.

To increase the depth of RNA-Seq sequencing and mapping rate of sequencing reads [19, 27], ribosomal RNA was removed by using the RiboZero rRNA removal kit (Epicentre) prior to cDNA library preparation. Double stranded cDNA library was prepared by using NEBNext® Ultra™ Directional RNA Library Prep kit for Illumina (New England Biolabs) following manufacturer's protocol from $1\mu\text{g}$ of total RNA. Kappa cDNA library quality control was performed prior to pooling libraries for flow cell amplification. All cDNA libraries were sequenced using Illumina HiSeq 2000 to produce 50 million reads, 50bp paired-end reads with multiplexing (4 samples/lane) (cDNA library preparation and RNA-Seq run was performed by Genomic Service Lab at Hudson Alpha).

RNA-Seq reads processing and alignment

TopHat2 [28] was used to map the sequencing reads to the human genome (UCSC hg19) by allowing multiple hits. Mapping rate was calculated as the percentage of read that were properly mapped. Samples with low mapping rates ($<20\%$) were discarded from further analysis. Cufflinks 2.2.0 [29] was used to calculate FPKMs values as the estimated gene expression levels. FPKMs between IPF and controls were compared using Cuffdiff and genes with a false discovery rate (FDR) adjusted $p < 0.05$ were identified as differentially expressed genes (DEGs) [30]. Cuffdiff test assigns a status of genes as: OK - test successful, NOTEST - not enough alignments for testing, LOWDATA - too complex or shallowly sequenced, HIDATA - too many fragments in locus, or FAIL - an ill-conditioned covariance matrix or other numerical exception prevents testing [31]. In this study we report genes with OK status as genes with sufficient coverage (Additional file 1: Table S1). Multidimensional scaling (MDS) analysis was performed for data visualization [32].

Microarray data

We used previously generated microarray data (Agilent). Briefly, the normalization of the gProcessed signal was performed using cyclic-LOESS and bioconductor package. Complete datasets and protocols were previously published [24, 26], and deposited in data repository GEO (accession no. GSE47460), and are also available in the Lung Genomics Research Consortium's (LGRC) website (<http://www.lung-genomics.org/>). Fold change and FDR values were calculated using the Significance Analysis of Microarrays (SAM) tools. Microarray experiments were compliant with MIAME guidelines.

Comparison of IPF signatures obtained from FFPE RNA-Seq and from microarray analysis of fresh frozen tissue

Considering that nearly all of the transcriptomic data in IPF was generated on fresh frozen tissues we compared

IPF signatures obtained with FFPE RNA-Seq (Illumina) with those obtained from Microarray (Agilent) gene expression data. Unique gene probes from microarrays ($n = 16,741$) were matched with genes having sufficient coverage by RNA-Seq ($n = 15,149$) by the gene symbol which provided a matched set of 13,304 genes. For analysis of differentially expressed genes, Significance Analysis of Microarrays (SAM) was used in the case of microarrays, and Cuffdiff was used in the case of RNA-Seq. Significance was defined as FDR adjusted $p < 0.05$. A discordant gene was defined as a significant gene that was increased in microarray but decreased in RNA-Seq and vice versa. The fold change (FC) of each gene was calculated by dividing the average mean value of IPF group by the average mean value of control group for both RNA-Seq and microarray. Log_2FC is the log base 2 transformed FC. Log_2FC of microarray were plotted against Log_2FC of RNA-Seq values for matching genes between the two platforms.

Pathway enrichment analyses

Pathway enrichment analysis of the common Differentially Expressed Genes (DEGs) between the two techniques, RNA-Seq and Microarrays, was performed using MetaCore (Thomson Reuters). In this way, we identified the top 13 statistically significant enriched pathways with FDR adjusted p value < 0.1 for pathways of common increased genes, and the top 50 statistically significant enriched pathways with FDR adjusted $p < 0.1$ for pathways of common decreased genes. Gene candidates with fold change > 1 from top pathways were presented in the heatmap to show the distribution of IPF relevant genes among the pathways. To analyze the interaction between members of the gene network in RNA-Seq and microarray data independently, networks were built around MMP7, one of the most-widely studied IPF-relevant genes [33]. When building the network, we used all RNA-Seq (4,131) and all microarray (5,859) differentially expressed genes. With auto-expand and canonical options, the network was built and drawn by plotting genes from datasets to pre-built MetaCore network for MMP7. Members of the gene networks with their expression values and interactions were compared.

NanoString nCounter® gene expression quantification and validation

100ng of RNA isolated from fresh frozen lung tissues and 250ng of RNA isolated from FFPE lung tissues, as suggested by the NanoString protocol, were used in experiment. 7 CTRL and 8 IPF FF samples, and 5 CTRL and 7 IPF FFPE samples were used for validation. Probe set for each gene were designed and synthesized by NanoString nCounter®. For validation, we focused on genes that had at least two-fold change in at least one

dataset. Thus we selected 10 discordant genes, 10 genes that were significantly differentially expressed only in one dataset and 15 concordant genes (Additional file 2: Table S5). The discordant genes were: IFNG, BCL11B, PDPR, ADAM33, VPS13B, TNRC18, PPBP, PHYHIP, KRT14, ITLN1. The differentially expressed genes in only one dataset were: MMP1, COL1A1, SERPIND1, ADAM23, COL1A2, MMP9, WISP1, WNT2, FOS, FGG. The concordant genes were: SERPINE1, CXCL2, MMP19, CAV1, CTNNA1, WNT10A, MMP7, SPP1, CXCL13, PLA2G2A, POSTN, LAMA3, COL17A1, SLC6A4, ANKRD1. We used RNA isolated from 7 IPF and 5 CTRL FFPE tissues and RNA isolated from 8 IPF and 7 CTRL fresh frozen tissues from the LGRC cohort. We have followed a standard manufacturer protocol for sample preparation, hybridization and detection. Data were analyzed using nSolver 3.0 digital analyzer software.

Results

Quality of RNA isolated from FFPE tissues

Consistent with previous reports [15, 25] the integrity of RNA isolated from FFPE tissues was decreased probably due to formalin fixation and archival time. All RNA isolations had $\text{OD}_{260/280} > 1.9$ confirming the high purity of RNA (data not shown). The RIN numbers were in the range of 2.1–2.6 and the most abundant RNA fragments were in the range of 100–150 ribonucleotides for all samples (Additional file 3: Figure S1) suggesting a similar degradation and quality regardless whether FFPE tissues were obtained from controls or IPF patients. Repeat isolations of RNA per FFPE block had very similar fragmentation and the same RIN number (Additional file 3: Figure S1). Archival time had no effect on RNA quality. We proceeded with one of the RNA isolation per FFPE block for cDNA library preparations and RNA-Seq analysis.

Mapping, transcript quantification and analysis of differentially expressed genes of FFPE RNA-Seq data

Mapping sequencing reads to the human genome produced an average of ~116 million reads at 50bp per sample with ~62 million mapped reads (mapping rate 48%). Only one RNA sample, corresponding to FFPE 7 (Additional file 3: Figure S1) had a lower mapping rate, 20.94% corresponding to 15.9 million reads (Table 1). This sample was excluded from downstream analysis. While we did not observe an effect of archival time on RIN number, samples with archival time longer than 7 years had lower mapping rates (40–50 million reads), a finding consistent with previous reports [21]. RNA-Seq identified 15,149 genes with sufficient coverage out of the 23,615 annotated genes in hg19, after filtering out genes that did not have enough alignments for testing, were too complex, had low number of sequencing reads

Table 1 Summary of RNA-Seq (FFPE) gene mapping

FFPE	RNA Seq	Orig Reads	QC failed reads	Unmapped Reads	Mapped Reads	Hits	Proper hits	Mapping Rate	Hits Rate	Proper hits Rate
1	SL32670	109112224	1585386	55426188	52100650	61090707	49057850	0.4775	0.5599	0.4496
2	SL32671	112024858	1901325	65686848	44436685	52541559	40745436	0.3967	0.469	0.3637
3	SL32674	137705852	1225400	58878848	77601604	93224772	74057010	0.5635	0.677	0.5378
4	SL32677	114532074	3536592	65895279	45100203	52390752	36239592	0.3938	0.4574	0.3164
5	SL32679	94880502	1404145	41223487	52252870	60762016	46203712	0.5507	0.6404	0.487
6	SL32680	98383774	3001192	45024767	50357815	58405722	44600578	0.5119	0.5937	0.4533
7	SL32681	76149832	2406659	57793748	15949425	17505889	8309682	0.2094	0.2299	0.1091
1C	SL32683	191953246	1025564	44983418	145944264	1.72E + 08	135808266	0.7603	0.8955	0.7075
2C	SL71047	126656896	4828921	72935707	48892268	59052405	40219390	0.386	0.4662	0.3175
3C	SL71048	92915968	5523822	45167979	42224167	51054887	38389264	0.4544	0.5495	0.4132
4C	SL71049	130059384	8904855	55527773	65626756	82086047	60494026	0.5046	0.6311	0.4651
5C	SL71050	121812940	7765072	43474914	70572954	83920938	60927332	0.5794	0.6889	0.5002

or had too many fragments in locus. Out of those 15,149 genes, Cuffdiff identified 4,131 differentially expressed genes (FDR < 0.05), including 1,920 increased genes and 2,211 decreased genes (Additional file 1: Table S1). Multidimensional scaling (MDS) analysis demonstrated a clear separation between control and IPF FFPE samples based on expression profiles (Fig. 2).

Comparison of differentially expressed genes of FFPE RNA-Seq to Microarray analysis of fresh frozen tissues

The IPF expression profile has been well defined by microarray analysis of RNA isolated from fresh frozen IPF

lung tissues [34–36]. To validate gene expression profiles obtained from FFPE IPF tissues, we compared it to gene expression data from microarrays based on RNA isolated from fresh frozen lungs (GSE47460, [24]). Figure 3a, depicts the comparison of microarray and RNA-Seq datasets. Microarray analysis demonstrated 2,306 increased and 3,367 decreased genes (FDR < 0.05) between IPF and controls while FFPE RNA-Seq identified 1,920 increased genes and 2,211 decreased genes. 760 increased and 1,413 decreased genes overlapped between microarrays and FFPE RNA-Seq (Fig. 3a, yellow and purple dots; Fig. 3b, Additional file 4: Table S4). Only 92 genes that were significantly differentially expressed were discordant between platforms (Fig. 3a, grey dots). 11,039 genes (Fig. 3a, white dots) were not differentially expressed (FDR > 0.05) in both datasets. 940 and 1,546 increased genes and 661 and 1,954 decreased genes that did not overlap between FFPE RNA-Seq and microarrays (Fig. 3b, white dots on Fig. 3a). To determine whether the overlap between the differentially expressed genes in FFPE RNA-Seq and FF microarrays in both datasets was not due to random association, we performed a hypergeometric test which revealed that the overlap was highly significant ($p < 10^{-182}$) for both increased and decreased genes. The hypergeometric test for discordant genes between FFPE RNA-Seq and FF microarrays revealed a probability of $p \sim 1$, suggesting that discordant genes are identified due to random association.

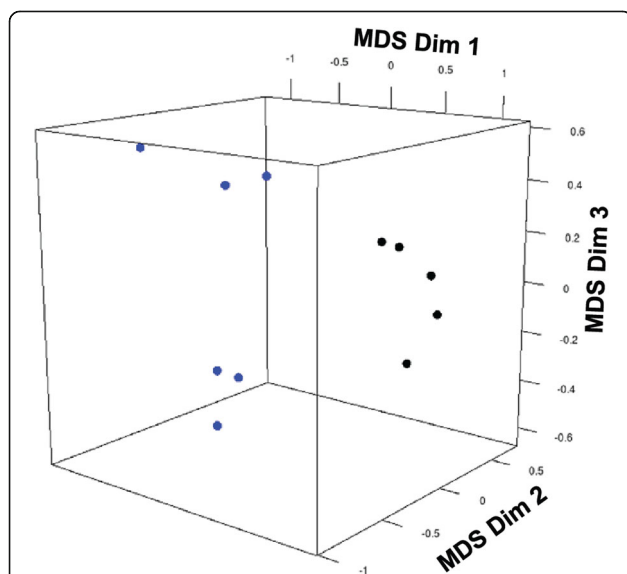
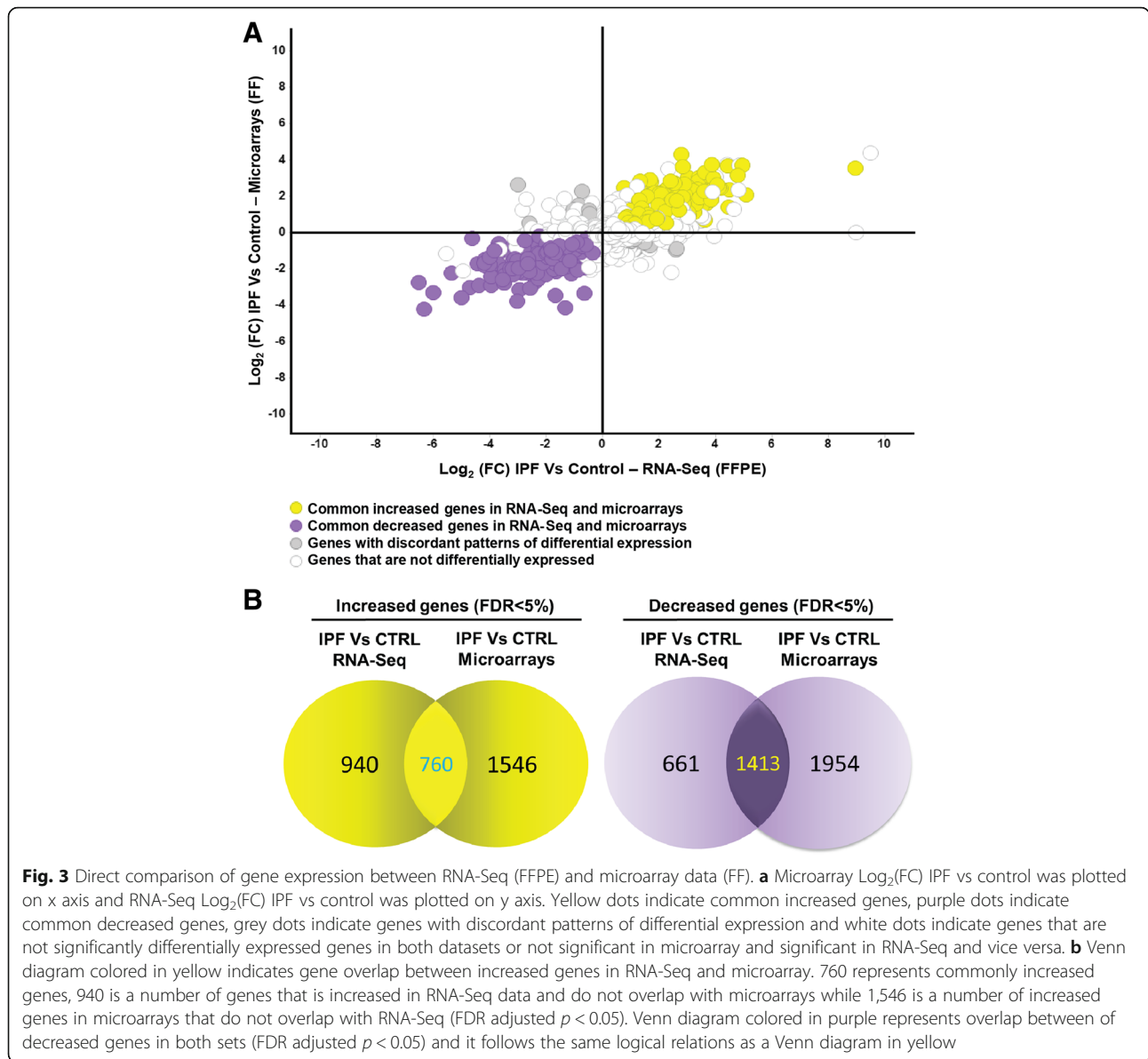


Fig. 2 MDS analysis based on gene expression demonstrates a clear separation between the IPF and control FFPE samples. The top three MDS dimensions based on the top 5,000 genes differentially expressed between IPF and control were plotted using edge R package for data visualization. Each dot is one sample. Blue represents IPF and black represents control, respectively

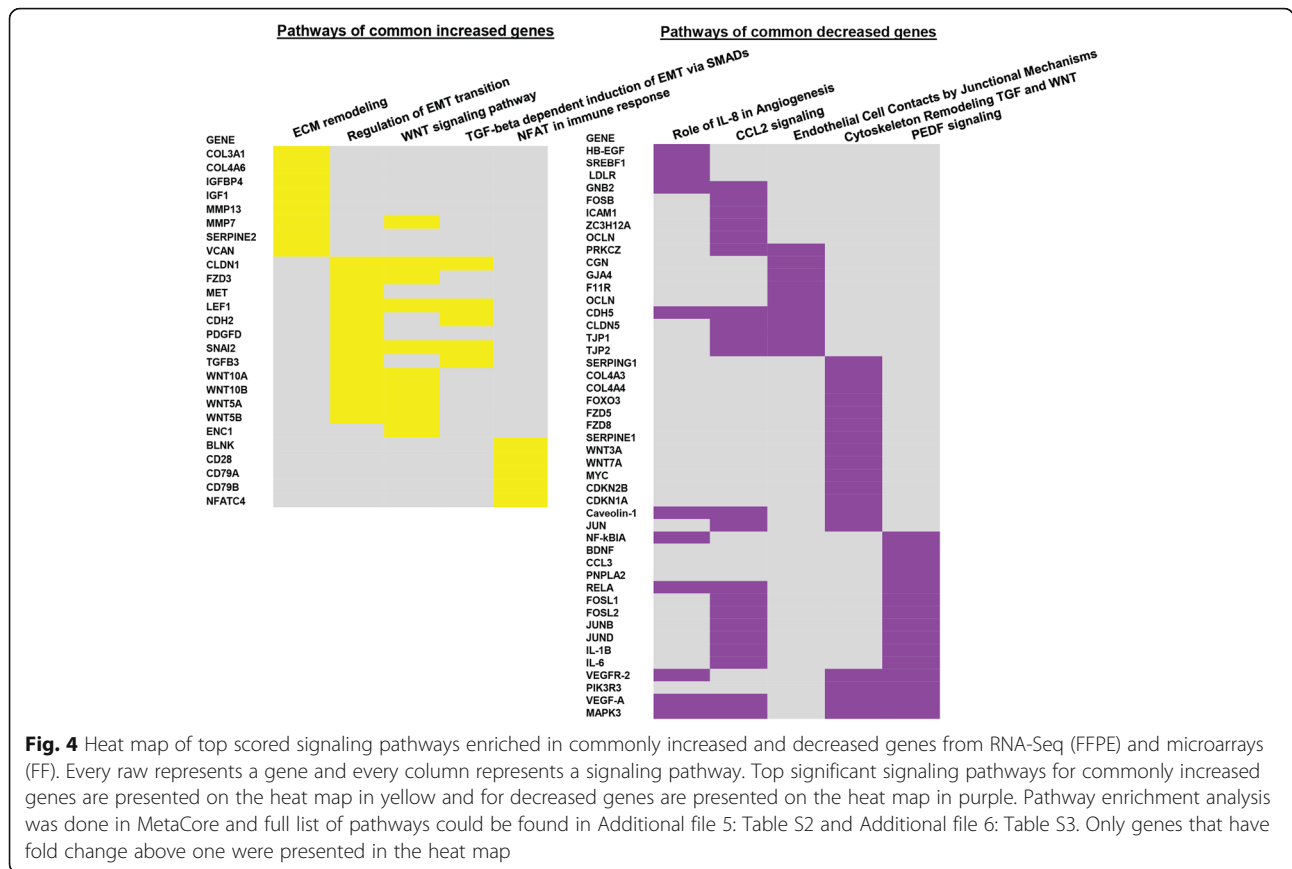
FFPE RNA-Seq results contain IPF relevant biological information

Previous studies on IPF gene expression profiles identified gene candidates and significant pathways directly involved in IPF development [34, 36–39]. To investigate whether IPF relevant genes and pathways could be detected in our RNA-Seq data, we analyzed canonical pathways and IPF relevant gene networks of RNA-Seq



and microarray overlapping genes by MetaCore (Fig. 4). Among the top 50 pathways significantly associated with our dataset (Additional file 5: Table S2 and Additional file 6: Table S3) we found many pathways known to play a significant role IPF [8, 37, 39] including developmental, cytoskeleton, extracellular remodeling and cell adhesion pathways. We performed a cluster analysis of genes that had a fold change above one and were present in at least one pathway. Figure 4 (left panel) represents a summary of increased genes in the following top five increased pathways: Extracellular matrix remodeling, regulation of EMT transition, WNT, TGF- β and NFAT pathways. Figure 4 (right panel) provides a summary of decreased genes in top five decreased pathways: IL8, endothelial

cell contacts, CCL2, cytoskeletal remodeling TGF/WNT, and PEDF signaling. We also identified several genes such as: COL3A1, COL4A6, MMP7 and MMP13, TGF-beta, WNT family, Serpine 1, LEF1, CLDN1 and CAV1 which had been shown to be relevant to IPF pathogenesis [33, 36, 40–44] suggesting that IPF relevant genes could be detected in FFPE tissues. To further support the notion that expression profiles from FFPE tissues are a valid source of information for transcriptomic profiling in IPE, we performed a network analysis for MMP7 gene, a well-known IPF relevant gene [24, 33, 45]. We hypothesized that building a network around MMP7 gene, should allow us to see if two datasets predict the same networking candidates and directions of interaction

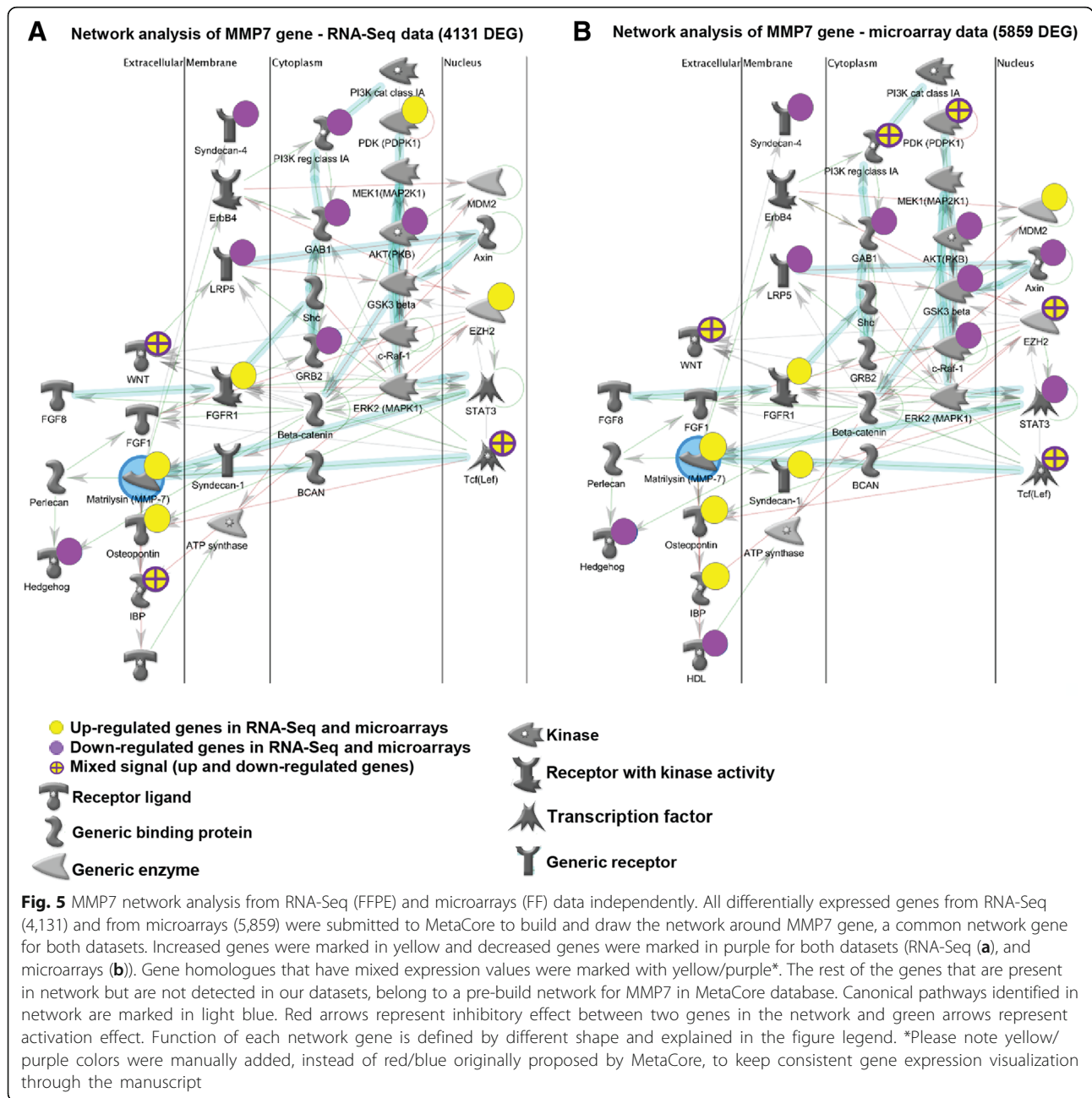


between candidates. For this purpose we performed an independent MetaCore network analyses using all RNA-Seq (4,131) and microarray (5,859) differentially expressed genes. Figure 5 shows that out of a total of 33 gene candidates proposed for RNA-Seq MMP7 network (Fig. 5a) and microarray MMP7 network (Fig. 5b), 15 genes were identified in RNA-Seq data and 21 genes were identified in microarray data. 14 genes overlapped between networks and 11 genes were not identified in any dataset. 7 genes in the MMP7 network (GSK3, c-Raf1, MDM2, Axin, Stat3, Syndecan 1, HDL) were differentially expressed in microarray data (with less than two fold change), but not in RNA-Seq data. Only one gene, GRB2, was differently expressed in RNA-Seq data (with less than two fold change) but not in microarray data. The 14 genes that overlap between two MMP7 networks represent 67% (14 out of 21) of all gene candidates for MMP7 network identified in our data and have preserved directions of interactions with surrounding genes. The hypergeometric test on 14 genes that overlap between microarray and RNA-Seq in network analysis revealed $p = 2.6 \times 10^{-30}$ for RNA-Seq and $p = 1.4 \times 10^{-28}$ for microarray confirming

that the gene overlap is highly significant. This demonstrates that FFPE data provides significant gene network information that is comparable to the gene network obtained from un-paired fresh tissues.

NanoString nCounter® gene expression validation in fresh frozen and FFPE tissues

We performed validation of gene expression levels by NanoString nCounter® technology. This technology performs better than RT-PCR in archived FFPE tissues [46]. Overall NanoString nCounter® results correlated well with both microarrays ($r = 0.92$) and RNA-Seq ($r = 0.90$). Detailed results for all genes are provided in Additional file 2: Table S5. The fold change directionality of all 15 concordant genes was confirmed by NanoString (Fig. 6a, b). Out of 10 discordant genes, only 4 genes (IFNG, ITLN1, PPBP, VPS13B) remind discordant after NanoString validation (Fig. 6b and Additional file 7: Figure S2, lower panel). Out of 10 tested genes that were significantly expressed in at least one dataset only 2 genes from microarray (SERPIND1, WNT2) and 1 gene from

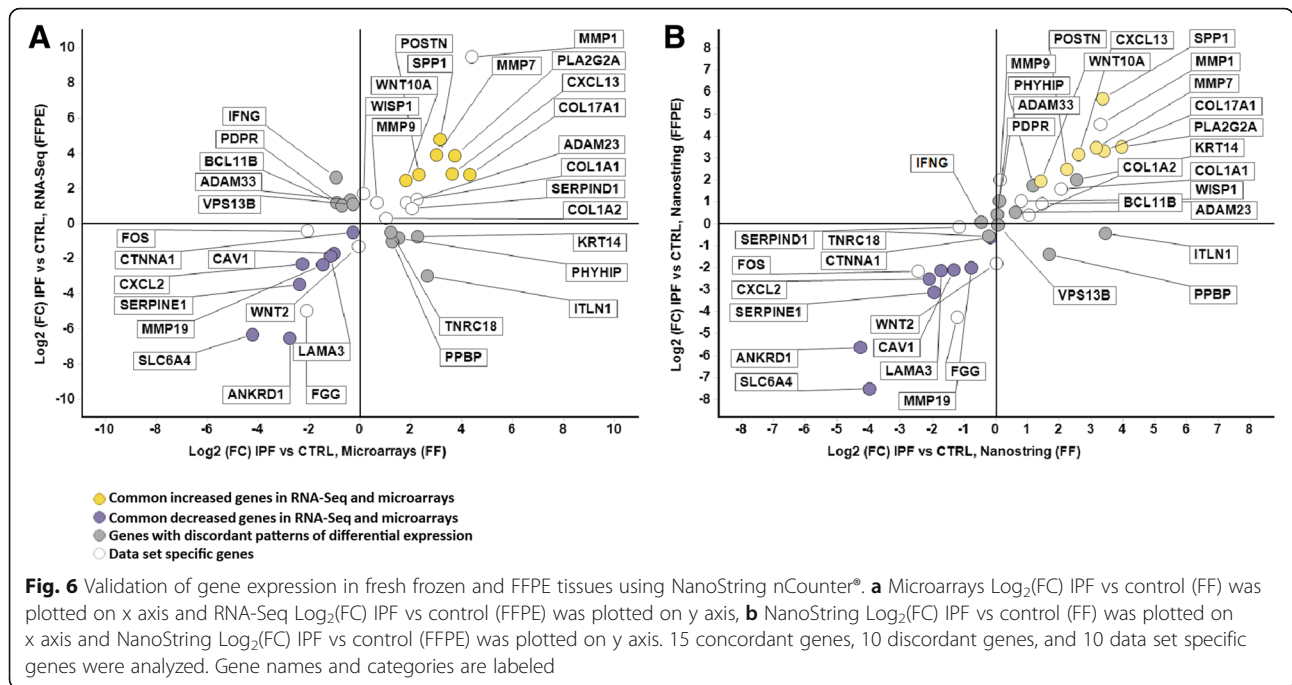


RNA-Seq (SERPIND1) were not confirmed (Fig. 6b, and Additional file 7: Figure S2, lower panels). Overall, we validated the significant changes for most genes for each type of samples (FF vs FFPE). This suggests that the source of discordance may have been tissue heterogeneity and samples being un-paired.

Discussion

Our study demonstrates that transcriptomic analysis of RNA isolated from FFPE IPF lung biopsies by RNA-Seq is feasible and the results comparable to

those obtained from gene expression microarrays. RNA-Seq resulted in an average of ~116 million 50 bp reads per sample with average of ~62 million mapped reads (Table 1). Our depth of sequencing, 62 million mapped reads, allowed for the detection of total of 15,149 genes with sufficient coverage, out of which 4,131 were differentially expressed between IPF and control (FDR adjusted $p < 0.05$). To validate the RNA-Seq FFPE results, we compared them to gene expression microarrays obtained from FF tissues and identified overlapping differentially expressed genes. The overlap was statistically significant.



The common genes were enriched for signaling pathways relevant to IPF such as: ECM remodeling process, WNT, TGF- β , NFAT, IL-8 in angiogenesis, CCL2 signaling and PEDF signaling [8, 36, 38, 39] and network analyses of both datasets revealed similar networks suggesting that FFPE RNA-Seq generated information that was relevant to IPF and comparable if not perfectly identical to FF tissue. Validation by NanoString nCounter® of concordant, discordant and dataset specific genes largely confirmed the results. Taken together, these findings demonstrate the feasibility and validity of RNA-Seq FFPE data and its relevance to IPF.

Although we provide strong evidence about the validity of our transcriptome RNA-Seq analysis, there are several discrepancies present when comparing RNA-Seq and microarray data. In addition to detecting the commonly expressed genes, we also detected differentially expressed genes that are discordant or do not overlap between RNA-Seq and microarray (Fig. 3b, gray and white dots on Fig. 3a, Additional file 4: Table S4). Computationally, we used MMP7 as a model gene to build the gene network and assess the potential bias in detecting gene interacting candidates from two datasets due to the presence of discordant or non-overlapping genes. For network analysis, we took into consideration all differentially expressed genes in RNA-Seq (4,131) and in microarray (5,859) independently. Out of 15 differentially

expressed genes from RNA-Seq and 21 differentially expressed genes from microarray that were found in the MMP7 network, 14 overlapped between datasets. 7 genes were only differentially expressed in microarray and one in RNA-Seq (Fig. 5). These 8 genes do not overlap between datasets (Fig. 3b, white dots on Fig. 3a) suggesting differences between datasets. To validate the results experimentally, and determine whether the results of comparison of different methodologies (microarrays vs RNA-Seq), or sample type (un-paired, FF vs FFPE) or other reasons we validated expression 35 genes: 10 discordant genes, 15 concordant genes, and 10 genes significant in only one dataset, using the NanoString nCounter®, a high precision system that measures gene expression based on digital color-coded barcode technology that provides significant accuracy and sensitivity and has been used successfully in partially degraded RNA samples. For most of the genes, directionality and significance of gene expression changes was confirmed. The number of discordant genes has been decreased with Nanostring nCounter® suggesting that differences in methodologies accounted for at least some of the differences.

The most significant limitations of our study are the small number of samples and the fact that we did not have paired FFPE and FF tissue from the same anatomical location in the same patient. Despite these limitations we found a significant overlap in differentially

expressed genes in both tissue types, a significant overlap in functional annotations of the functional annotations of these genes, and good correlation between our Nanostring nCounter® validation with either microarray analysis of FF tissue or RNA-Seq analysis of FFPE tissue. Our results are in agreement with recent observations that RNA-Seq analysis of FFPE tissues can generate valid and comparable gene expression to FF tissues [16, 20, 27]. In cancer tissues high correlation was observed between RNA-Seq of paired FF and FFPE tissues [19–22, 27, 47]. Important to note, that in all of those studies authors mention that while the information derived from FF or FFPE tissue is comparable, comparisons should be limited within one type of tissue processing. Direct comparison of FFPE diseased tissues to control FF tissues for example, would be highly confounded and probably generate spurious results.

The RNA we isolated from FFPE tissues was partially degraded as previously observed [13–17, 19, 25]. However, while we used FFPE biopsies that had a wide range of archival times, and handling performed at different hospitals, we did not find systematic differences between the tissues. Of 12 FFPE biopsies, we experienced low number of original reads and low mapping rate of only one FFPE sample was indistinguishable from other samples. Archival age, RNA and cDNA quality were similar to the other samples. This could be the result of conditions that cannot be directly observed such as minor changes in fixation technique, or storage that could induce changes in the RNA structures [21]. Thus, in our relatively small study the technical success of RNA-Seq from FFPE tissue was 92%. It is plausible, that in prospectively designed studies and standardized fixation protocols the success rate would be even higher [48].

To the best of our knowledge, our study is the first to demonstrate the feasibility of RNA-Seq of FFPE IPF lung samples. We hope and believe that the availability of our protocols, as well as our results, will facilitate the use of FFPE tissue for genome scale transcript profiling of IPF. This will overcome the limitation on availability of FF tissues and increase the capacity for transcriptomic profiling of IPF [47, 49, 50].

Conclusion

Our study serves as a proof of principle, that RNA-Seq performed on RNA isolated from archival FFPE IPF lung tissues is feasible, and reveals a gene expression profile relevant for IPF. This study further shows that there is a high concordance between RNA-Seq (FFPE) and microarray (FF) expression profiles for biopsies performed on different patients, and at different hospitals encouraging the further usage of FFPE biopsies. Taking into consideration the great potential

for transcriptomic research, FFPE tissues should be considered to overcome limitations in the availability of FF human lung tissues.

Additional files

Additional file 1: Table S1. FPKM p-value matrices. (XLS 15268 kb)

Additional file 2: Table S5. NanoString validation. (XLS 71 kb)

Additional file 3: Figure S1. Quality of RNA isolated from control and IPF FFPE lung tissues. To analyze RNA isolated from five 10- μ m slices of the whole FFPE lung tissue, RNA samples were run on the 2100 Bioanalyzer. RNA ladder was run to determine the size of RNA fragments. Two isolations of RNA per FFPE block, for control and IPF lung tissues, were performed. RIN numbers are presented at the panels. (PNG 632 kb)

Additional file 4: Table S4. IPF vs CTRL, RNA-Seq vs Microarrays, uploaded to Spotfire. (XLS 23589 kb)

Additional file 5: Table S2. Enrichment pathway analysis MetaCore_common decreased genes_microarrays vs RNA-Seq. (XLS 68 kb)

Additional file 6: Table S3. Enrichment pathway analysis MetaCore_common increased genes_microarrays vs RNA-Seq. (XLS 55 kb)

Additional file 7: Figure S2. Detailed results of NanoString nCounter validation. Upper 3 panels. Microarrays $\text{Log}_2(\text{FC})$ IPF vs control (FF) was plotted on x axis and RNA-Seq $\text{Log}_2(\text{FC})$ IPF vs control (FFPE) was plotted on y axis, than Nanostring $\text{Log}_2(\text{FC})$ IPF vs control (FF) was plotted on y axis and Microarrays $\text{Log}_2(\text{FC})$ IPF vs control (FF) was plotted on x axis, than Nanostring $\text{Log}_2(\text{FC})$ IPF vs control (FFPE) was plotted on y axis and RNA-Seq $\text{Log}_2(\text{FC})$ IPF vs control (FFPE) was plotted on x axis. 15 concordant genes between microarray and RNA-Seq were validated with NanoString. Gene names are labeled. FDR and p values are also labeled for each gene and technology. Similar to above, middle 3 panels present 10 discordant genes, and lower 3 panels present 10 genes in specific data set. (PNG 372 kb)

Abbreviations

DEG: Differentially expressed gene; FC: Fold change; FFPE: Formalin fixed paraffin embedded; FPKM: Fragments per kilobase of exon per million; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; IRB: Institutional Review Board; LGRC: The lung genomic research consortium; MDS: Multidimensional scaling; MIAME: Minimum information about a microarray experiment; NGS: Next generation sequencing; OD: Optical density; RIN: RNA Integrity number; RNA-Seq: RNA sequencing; RNA: Ribonucleic acid; rRNA: Ribosomal RNA; SAM: Significance analysis of microarrays; UIP: Usual interstitial pneumonia

Acknowledgements

RNA quality assessment, cDNA library preparation and RNA-Seq run were performed by Genomic Service Lab at Hudson Alpha. We thank Cynthia Vied and Michelle Arbeitman from College of Medicine, Florida State University, for valuable discussions regarding the study design and RNA-Seq.

Funding

Work was supported in part by National Institutes of Health grant R01HL127349 to NK, Harold Amos Faculty development program of the Robert Wood Johnson Foundation Award and the Pulmonary Fibrosis Foundation to JHM and the FSU College of Medicine Internal Seed Grant to BS, JB, MV.

Availability of data and materials

The microarray datasets supporting the conclusions of this article are available in the GEO repository (accession no. GSE47460 and <http://www.lung-genomics.org>).

The RNA-Seq dataset, representing normalized gene expression values (FPKM), and RNA-Seq vs Microarray dataset, representing comparison of gene

expression values, supporting the conclusions of this article are included within the article supplement files and uploaded to GEO under GSE83717.

Authors' contributions

MV conceived the study, participated in study design and coordination, data collection and analysis, drafted the manuscript, JHM participated in study design, data collection and analysis, helped to draft the manuscript, JB helped conceive the study, participated in study design and histopathological evaluation of FFPE tissues, VST helped conceive the study, participated in FFPE tissues collection, clinical evaluation of IPF patients and study design, DJ participated in FFPE tissues collection, clinical evaluation of IPF patients and study design, VZ participated in literature review, helped with FFPE tissues collection, study design and coordination, SP participated in data analysis and valuable discussions, JS and RH participated in FFPE tissues collection and histopathological analysis, XY participated in RNA-seq data analysis, BS conceived the study, participated in study design, data collection and analysis and helped to draft the manuscript, NK participated in study design, generation of microarray and NanoString nCounter data, led data analysis and drafted the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

NK consulted Biogen Idec, Boehringer Ingelheim, Third Rock, MMI, and Pliant and received a grant from Biogen Idec and non-financial support from with MiRagen, all outside the submitted work. In addition, NK has a patent New Therapies in Pulmonary Fibrosis licensed to Quitsa/SLI, and a patent on Peripheral Blood Gene Expression as diagnostic in Pulmonary Fibrosis. All of NK's competing interests are outside the submitted work. All other authors declare no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

This study used Lung FFPE biopsies from the departmental FFPE archives of the Clinic for Pulmonology in Belgrade ($n = 8$), and the NHLBI funded Lung Tissue Research Consortium ($n = 4$) as previously described [51]. The use of samples was approved by the Institutional Review Boards (IRB) at the Clinical Center of Serbia (approval number 4072/54), University of Pittsburgh (approval number IRB0411036) and Yale School of Medicine (approval number 1409014689). Informed consents were obtained as appropriate according to IRB.

Author details

¹Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL, USA. ²Faculty of Medicine, University of Belgrade, Belgrade, Serbia. ³Clinic for Pulmonology, Clinical Center of Serbia, Belgrade, Serbia. ⁴Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia. ⁵Department of Thoracopulmonary Pathology, Service of Pathology, Clinical Centre of Serbia, Belgrade, Serbia. ⁶Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia. ⁷Section of Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine, New Haven, CT, USA. ⁸Department of Pathology, Yale University School of Medicine, New Haven, CT, USA. ⁹Pathology and Laboratory Medicine Service, VA CT Healthcare System, West Haven, CT, USA.

Received: 14 April 2016 Accepted: 20 December 2016

Published online: 12 January 2017

References

- Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and Prevalence of Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2006;174(7):810–6.
- King Jr TE, Pardo A, Selman M. Idiopathic pulmonary fibrosis. *Lancet*. 2011; 378(9807):1949–61.
- Selman M, Pardo A. Idiopathic pulmonary fibrosis: an epithelial/fibroblastic cross-talk disorder. *Respir Res*. 2002;3:3.
- Kass DJ, Kaminski N. Evolving genomic approaches to idiopathic pulmonary fibrosis: moving beyond genes. *Clin Transl Sci*. 2011;4(5):372–9.
- Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier J-F, Flaherty KR, Lasky JA, et al. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *Am J Respir Crit Care Med*. 2011;183(6):788–824.
- Smith M, Dalurzo M, Panse P, Parish J, Leslie K. Usual interstitial pneumonia-pattern fibrosis in surgical lung biopsies. Clinical, radiological and histopathological clues to aetiology. *J Clin Pathol*. 2013;66(10):896–903.
- Kaarteenaho R. The current position of surgical lung biopsy in the diagnosis of idiopathic pulmonary fibrosis. *Respir Res*. 2013;14:43.
- Herazo-Maya JD, Kaminski N. Personalized medicine: applying 'omics' to lung fibrosis. *Biomark Med*. 2012;6(4):529–40.
- Kim SY, Diggans J, Pankratz D, Huang J, Pagan M, Sindy N, Tom E, Anderson J, Choi Y, Lynch DA, et al. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Res Med*. 2015;3(6):473–82.
- DePianto DJ, Chandriani S, Abbas AR, Jia G, N'Diaye EN, Caplazi P, Kauder SE, Biswas S, Karnik SK, Ha C, et al. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax*. 2015;70(1):48–56.
- Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, Rosen R, Neidermyer AJ, McKean DF, Groshong SD, et al. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax*. 2013;68(12):1114–21.
- Maher TM. Transcriptional phenotyping of fibrotic lung disease: a new gold standard? *Lancet Res Med*. 2015;3(6):423–4.
- Farragher SM, Tanney A, Kennedy RD, Paul Harkin D. RNA expression analysis from formalin fixed paraffin embedded tissues. *Histochem Cell Biol*. 2008;130(3):435–45.
- Gilbert MT, Haselkorn T, Bunce M, Sanchez JJ, Lucas SB, Jewell LD, Van Marck E, Worobey M. The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? *PLoS One*. 2007;2(6):e537.
- Bibikova M, Talantov D, Chudin E, Yeakley JM, Chen J, Doucet D, Wickham E, Atkins D, Barker D, Chee M, et al. Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am J Pathol*. 2004;165(5):1799–807.
- Frank M, Döring C, Metzler D, Eckerle S, Hansmann M-L. Global gene expression profiling of formalin-fixed paraffin-embedded tumor samples: a comparison to snap-frozen material using oligonucleotide microarrays. *Virchows Arch*. 2007;450(6):699–711.
- Hosey AM, Gorski JJ, Murray MM, Quinn JE, Chung WY, Stewart GE, James CR, Farragher SM, Mulligan JM, Scott AN, et al. Molecular basis for estrogen receptor alpha deficiency in BRCA1-linked breast cancer. *J Natl Cancer Inst*. 2007;99(22):1683–94.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63.
- Liu Y, Noon AP, Aguiar Cabeza E, Shen J, Kuk C, Ilczynski C, Ni R, Sukhu B, Chan K, Barbosa-Morais NL, et al. Next-generation RNA Sequencing of Archival Formalin-fixed Paraffin-embedded Urothelial Bladder Cancer. *Eur Urol*. 2014.
- Li P, Conley A, Zhang H, Kim H. Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq. *BMC Genomics*. 2014;15(1):1087.
- Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtroder K, Kruhoffer M, Hager H, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One*. 2014;9(5):e98187.
- Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, Taubenberger JK. High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *J Pathol*. 2013;229(4):535–45.
- Sinicropi D, Qu K, Collin F, Crager M, Liu M, Pelham R, Pho M, Dei Rossi A, Jeong J, Scott A, et al. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. *PLoS One*. 2012;7:e40092.
- Bauer Y, Tedrow J, de Bernard S, Birker-Robaczewska M, Gibson KF, Guardela BJ, Hess P, Klenk A, Lindell KO, Poirey S, et al. A Novel Genomic Signature with Translational Significance for Human Idiopathic Pulmonary Fibrosis. *Am J Respir Cell Mol Biol*. 2014;52(2):217–31.
- Glenn ST, Jones CA, Liang P, Kaushik D, Gross KW, Kim HL. Expression profiling of archival renal tumors by quantitative PCR to validate prognostic markers. *Biotechniques*. 2007;43(5):639–40. 642–633, 647.

26. Kim S, Herazo-Maya JD, Kang DD, Juan-Guardela BM, Tedrow J, Martinez FJ, Sciruba FC, Tseng GC, Kaminski N. Integrative phenotyping framework (IPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics*. 2015;16:924.
27. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*. 2014;15(1):419.
28. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36.
29. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
30. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q -value. *Ann Stat*. 2003;31(6):2013–35.
31. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562–78.
32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*. 2010;26(1):139–40.
33. Rosas IO, Richards TJ, Konishi K, Zhang Y, Gibson K, Lokshin AE, Lindell KO, Cisneros J, Macdonald SD, Pardo A, et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med*. 2008;5(4).
34. Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJD, Morris D, Huang X, Sheppard D, Heller RA. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci*. 2000;97(4):1778–83.
35. Kaminski N. Microarray analysis of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol*. 2003;29(3 Suppl):S32–36.
36. Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, Aziz N, Kaminski N, Zlotnik A. Gene Expression Profiles Distinguish Idiopathic Pulmonary Fibrosis from Hypersensitivity Pneumonitis. *Am J Respir Crit Care Med*. 2006;173(2):188–98.
37. Boon K, Bailey NW, Yang J, Steel MP, Groshong S, Kervitsky D, Brown KK, Schwarz MI, Schwartz DA. Molecular phenotypes distinguish patients with relatively stable from progressive idiopathic pulmonary fibrosis (IPF). *PLoS One*. 2009;4(4):e5134.
38. Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW, et al. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2009;180(2):167–75.
39. Selman M, Pardo A, Kaminski N. Idiopathic pulmonary fibrosis: aberrant recapitulation of developmental programs? *PLoS Med*. 2008;5(3):e62.
40. Konigshoff M, Balsara N, Pfaff EM, Kramer M, Chrobak I, Seeger W, Eickelberg O. Functional Wnt signaling is increased in idiopathic pulmonary fibrosis. *PLoS One*. 2008;3(5):e2142.
41. Vuga LJ, Ben-Yehudah A, Kovkarova-Naumovski E, Oriss T, Gibson KF, Feghali-Bostwick C, Kaminski N. WNT5A is a regulator of fibroblast proliferation and resistance to apoptosis. *Am J Respir Cell Mol Biol*. 2009;41(5):583–9.
42. Liu RM. Oxidative stress, plasminogen activator inhibitor 1, and lung fibrosis. *Antioxid Redox Signal*. 2008;10(2):303–19.
43. Wang XM, Zhang Y, Kim HP, Zhou Z, Feghali-Bostwick CA, Liu F, Ifedigbo E, Xu X, Oury TD, Kaminski N, et al. Caveolin-1: a critical regulator of lung fibrosis in idiopathic pulmonary fibrosis. *J Exp Med*. 2006;203(13):2895–906.
44. Lappi-Blanco E, Lehtonen ST, Sormunen R, Merikallio HM, Soini Y, Kaarteenaho RL. Divergence of tight and adherens junction factors in alveolar epithelium in pulmonary fibrosis. *Hum Pathol*. 2013;44(5):895–907.
45. Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, Lollini L, Morris D, Kim Y, DeLustro B, et al. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A*. 2002;99(9):6292–7.
46. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordóñez B, Gullane P, Irish J, Jurisica I, Kamel-Reid S. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol*. 2011;11:46.
47. Ley B, Brown KK, Collard HR. Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2014;307(9):L681–91.
48. Waldron L, Simpson P, Parmigiani G, Huttenhower C. Report on emerging technologies for translational bioinformatics: a symposium on gene expression profiling for archival tissues. *BMC Cancer*. 2012;12:124.
49. Zhang Y, Kaminski N. Biomarkers in idiopathic pulmonary fibrosis. *Curr Opin Pulm Med*. 2012;18(5):441–6.
50. Borensztajn K, Crestani B, Kolb M. Idiopathic pulmonary fibrosis: from epithelial injury to biomarkers—insights from the bench side. *Respiration*. 2013;86(6):441–52.
51. Kusko RL, Brothers 2nd JF, Tedrow J, Pandit K, Huleihel L, Perdomo C, Liu G, Juan-Guardela B, Kass D, Zhang S, 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(8):948–60.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

