Original Article

Plasma RNA profiling unveils transcriptional signatures associated with resistance to osimertinib in EGFR T790M positive non-small cell lung cancer patients

Andrey Alexeyenko[1,2,3], Odd Terje Brustugun[4,5,6], Inger Johanne Zwicky Eide[4,5,6], Radosveta Gencheva[7], Zeinab Kosibaty[7], Yi Lai[7], Luigi de Petris[7,8], Georgios Tsakonas[7,8], Oscar Grundberg[8], Bo Franzen[7], Kristina Viktorsson[7], Rolf Lewensohn[7,8], Per Hydbring[7*], Simon Ekman[7,8*]

[1]Science for Life Laboratory, Box 1031, Solna, Sweden; [2]Evi-networks consulting, Huddinge, Sweden; [3]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Solna, Sweden; [4]Section of Oncology, Drammen Hospital, Vestre Viken Hospital Trust, Drammen, Norway; [5]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; [6]Department of Cancer Genetics, Institute for Cancer Research, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway; [7]Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden; [8]Thoracic Oncology Center, Karolinska University Hospital, Stockholm, Sweden

Contributions: (I) Conception and design: A Alexeyenko, P Hydbring, S Ekman; (II) Administrative support: P Hydbring, S Ekman; (III) Provision of study materials or patients: IJZ Eide, R Lewensohn, G Tsakonas, O Grundberg, L de Petris, OT Brustugun, S Ekman; (IV) Collection and assembly of data: A Alexeyenko, IJZ Eide, R Gencheva; (V) Data analysis and interpretation: A Alexeyenko, P Hydbring, S Ekman; (VI) Manuscript writing: All authors. (VII) Final approval of manuscript: All authors.

*These authors contributed equally for the senior authorship.

Correspondence to: Dr. Per Hydbring. Akademiska Stråket 1, BioClinicum J6:20, 17164 Solna, Sweden. Email: per.hydbring@ki.se.

Background: Targeted therapy with tyrosine kinases inhibitors (TKIs) against epidermal growth factor receptor (EGFR) is part of routine clinical practice for EGFR mutant advanced non-small cell lung cancer (NSCLC) patients. These patients eventually develop resistance, frequently accompanied by a gatekeeper mutation, T790M. Osimertinib is a third-generation EGFR TKI displaying potency to the T790M resistance mutation. Here we aimed to analyze if exosomal RNAs, isolated from longitudinally sampled plasma of osimertinib-treated EGFR T790M NSCLC patients, could provide biomarkers of acquired resistance to osimertinib.

Methods: Plasma was collected at baseline and progression of disease from 20 patients treated with osimertinib in the multicenter phase II study TKI in Relapsed EGFR-mutated non-small cell lung cancer patients (TREM). Plasma was centrifuged at 16,000 g followed by exosomal RNA extraction using Qiagen exoRNeasy kit. RNA was subjected to transcriptomics analysis with Clariom D.

Results: Transcriptome profiling revealed differential expression [log2(fold-change) >0.25, false discovery rate (FDR) P<0.15, and P(interaction) >0.05] of 128 transcripts. We applied network enrichment analysis (NEA) at the pathway level in a large collection of functional gene sets. This overall enrichment analysis revealed alterations in pathways related to EGFR and PI3K as well as to syndecan and glypican pathways (NEA FDR $<3\times10^{-10}$). When applied to the 40 individual, sample-specific gene sets, the NEA detected 16 immune-related gene sets (FDR <0.25, P(interaction) >0.05 and NEA z-score exceeding 3 in at least one sample).

Conclusions: Our study demonstrates a potential usability of plasma-derived exosomal RNAs to characterize molecular phenotypes of emerging osimertinib resistance. Furthermore, it highlights the involvement of multiple RNA species in shaping the transcriptome landscape of osimertinib-refractory NSCLC patients.

Keywords: Non-small cell lung cancer (NSCLC); exosomal RNA; osimertinib; transcriptome; epidermal growth factor receptor (EGFR)

© Translational Lung Cancer Research. All rights reserved.

Submitted Mar 25, 2022. Accepted for publication Aug 22, 2022.
doi: 10.21037/tlcr-22-236
View this article at: https://dx.doi.org/10.21037/tlcr-22-236

Introduction

Lung cancer is the most common cancer worldwide resulting in nearly 20% of all cancer deaths (1). The most common subgroup of lung cancer is non-small cell lung cancer (NSCLC) constituting around 85% of lung cancer cases. Activating mutations in the gene encoding epidermal growth factor receptor (EGFR) occur in approximately 15% of all NSCLC adenocarcinomas in Western patients with the prevalence being 3–4 times higher in Asians (2). These mutations result in a constitutively active EGFR receptor promoting uncontrolled proliferation, invasion and metastasis (3). The vast majority of NSCLCs harboring activating mutations in EGFR respond favorably to first-generation ATP-competitive tyrosine kinase inhibitors (TKIs) erlotinib and gefitinib (4-6), or to second-generation EGFR TKIs afatinib and dacomitinib that irreversibly bind to the kinase domain. Although first- and second-generation TKIs are clinically favorable compared to platinum-based therapy, inevitably all such tumors develop resistance to these TKIs, which is partly a consequence of the emergence of a secondary mutation, T790M (7). This mutation results in an increased affinity for ATP, negating the efficacy of ATP-competitive TKIs (8). Osimertinib is a covalent irreversible third-generation TKI, which targets NSCLCs with activating mutations in EGFR regardless of presence of the T790M mutation (9). Osimertinib received U.S. Food and Drug Administration (FDA)-approval in 2017 as second-line therapy, followed by approval as first-line therapy in 2018 by both the FDA and the European Medicines Agency (EMA) (10-14). Despite the impressive effects of osimertinib in the clinical setting, patients receiving osimertinib eventually develop resistance. Known resistance mechanisms involve acquired mutation of the drug-binding cysteine, C797S, as well as amplifications of *MET*, *HER2* and *PIK3CA*, together accounting for up to 50% of resistant cases. Furthermore, a large fraction of EGFR T790M NSCLCs progressing on osimertinib exhibit lost T790M-status (15-17). Several of the reported resistance mechanisms impact cell signaling pathways, likely altering gene transcription programs. Therefore, investigation of transcriptional changes is imperative for uncovering RNA biomarkers as well as understanding potential mechanisms and biological outcomes of acquired resistance to osimertinib. This requires repeated biopsies to capture the progression of the disease. Repeated solid tissue biopsy sampling presents an invasive clinical procedure that only captures the molecular nature of the cells at the sampling site. In contrast, blood liquid biopsies will potentially capture all RNAs shed into the bloodstream in extracellular vesicles, including exosomes, from any tumor cell, potentially minimizing tumor heterogeneity. The drawback of liquid biopsies is that the subsequent profiling cannot distinguish tumor-derived vesicle-bound RNA from vesicle-bound RNA shed from healthy tissues. However, studying datasets of tens or even hundreds of solid tissue biopsy samples does not guarantee the identification of a common denominator for a specific phenotype. This demands rigorous cross-validation approaches, involving independently collected datasets. Furthermore, analysis would gain power by summarizing sparse individual gene events to the pathway level. The method of network enrichment analysis applied here allowed accounting for any altered transcripts regardless of expression of pathway genes (18,19). Using this approach, in an aim to identify potential new biomarkers of resistance, we have studied longitudinal changes in the transcriptional landscape of plasma extracellular vesicle bound RNAs from a cohort of EGFR T790M NSCLC patients receiving osimertinib as second-line treatment, and demonstrate that blood plasma serves as a comprehensive RNA source and that our major biological findings from plasma extracellular vesicle bound RNAs are corroborated using public and newly generated cell model data. We present the following article in accordance with the MDAR reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-236/rc).

Methods

Patient cohort and sample preparation

Twenty patients were included in the study. Five patients were males (38, 68, 43, 65 and 78 years old) and fifteen patients were females (62, 69, 53, 73, 56, 60, 75, 79, 59, 75, 76, 66, 69, 61 and 64 years old). All patients were enrolled

© Translational Lung Cancer Research. All rights reserved.

in the Northern European multicenter phase II TREM study (EudraCT No. 2015-000307-10) and diagnosed with EGFR T790M-mutant NSCLC with a treatment history involving disease progression on minimum one first and/or second-generation EGFR TKI (20). The patients were treated with osimertinib 80 mg daily until radiologic progression. In this osimertinib-treated cohort, samples of twelve patients were collected via the Oslo University Hospital and eight patients at the Karolinska University Hospital. Whole blood was drawn just before treatment start (baseline) and at radiological disease progression on osimertinib treatment while patients were still on osimertinib and just before change of therapy. Plasma was separated through centrifugal isolation, 2,000 g for 15 min, and aliquoted to fresh 1ml tubes and stored at -80C. The regional ethical committees at respective hospitals approved sampling for this study.

EGFR mutant parental cell lines NCI-H1975 and HCC827 and TKI-refractory cell lines (erlotinib-resistant HCC827, gefitinib-resistant HCC827, osimertinib-resistant HCC827 and osimertinib-resistant NCI-H1975) (21) were cultured in RPMI-1640 medium with 10% supplemented Fetal Bovine Serum at 5% CO_2, 37 ℃, and passaged when reaching sub-confluent conditions.

RNA extraction

Exosomal RNA was isolated at Karolinska Institutet. 1 mL plasma/sample point was centrifuged at 16,000 g for 10 minutes followed by processing using the ExoRNeasy serum plasma midi kit (Qiagen), as previously described (22), and the RNA was eluted in 14 µL RNase free water. RNA quantity and quality were assessed through the documentation of RNA integrity number (RIN) curves. All samples selected for analysis displayed similar RIN curves with a range from 1.50–2.90. Cell line total RNA was extracted from EGFR mutant TKI-refractory NSCLC cell lines (erlotinib-resistant HCC827, gefitinib-resistant HCC827, osimertinib-resistant HCC827 and osimertinib-resistant NCI-H1975) and EGFR mutant parental NSCLC cell lines (HCC827 and NCI-H1975) (21) using mirVana miRNA isolation kit (ThermoFisher Scientific Cat #AM1560).

Transcriptome analysis

3 µL of eluted exosomal total RNA, or cell line total RNA, was pre-amplified for 6 cycles before loaded onto Clariom D Pico Assay, human transcriptome arrays (ThermoFisher Scientific #902925). Cell line total RNA was loaded onto Clariom D Pico Assay in biological duplicates. Transcript expression values were normalized using Signal Space Transformation (SST-RMA) method.

Exploratory and statistical analyses

The SST-RMA values were bell-shape distributed, although the right tail was too extended. Therefore, the values were further log-transformed in order to render distribution closer to Gaussian and ensure homoscedasticity and usage of parametric statistics. However, the fold change values and boxplots visualization were based on the original SST-RMA values. Principal component analysis, Volcano plot analysis, RNA-class distribution analysis and differential gene expression analysis was performed for 81042 transcripts with Clariome annotation in R environment using functions from package base. The removal of batch effects, generalized least squared models and network enrichment analysis were implemented with R packages limma, nlme and NEArender, respectively (https://cran.r-project.org/web/packages/).

Network enrichment analysis

The network enrichment analysis (NEA) employs the global network, which combines all major types of molecular interactions in an unbiased way. By utilizing this topological information, NEA can render experimentally observed molecular alterations into a space of pathways and processes. The pathway view enables lower dimensionality, is more transparent for biological interpretation compared to other multivariate methods, and is also more efficient in absence of replicates—which is a typical situation in patient sample collections.

Similarly to the well-known over-representation analysis (ORA), NEA can analyze experimental altered gene sets (AGS), such as top *N* differentially expressed genes (DEG). AGSs are then tested for enrichment, i.e., significant "relatedness" with regard to (usually a large collection of) functional gene sets (FGS), such as pathways or other custom sets of biological importance. In ORA, enrichment of FGS versus AGS is determined by the fraction of genes shared by the two sets. NEA instead counts network edges that connect genes of AGS with any genes of FGS, which number is compared to a number expected by chance. The latter is influenced by variability in network edge numbers of involved genes and is therefore normalized by topological properties

© Translational Lung Cancer Research. All rights reserved.

(node degrees) of gene nodes in the global network.

NEA assigns profiles of FGS enrichment scores to each submitted AGS, which then could be used as either descriptive or predictive variables in the same way as gene expression profiles. NEA possesses higher statistical power to detect enrichment compared to ORA (18) and better reproducibility compared to both using raw gene profiles and diverse enrichment methods, which were tested in a systematic benchmark (19). Furthermore, NEA can identify network enrichment against e.g., signaling pathways, members of which not necessarily changed their own expression.

The integrated framework of NEA consists of three components: AGSs for each clinical sample, a sufficiently large collection of FGSs, and a version of global interaction network. The analysis was run in R environment using package NEArender of version 1.4. NEArender produced network enrichment scores for each AGS-FGS pair.

Global network version

The global network for NEA was a set of functional links from curated databases collected in the Pathway Commons project (version 9) (23) with 846,631 unique edges between 20,063 unique human gene nodes.

Functional gene sets (FGS)

The collection of pathways and gene sets included all entries from BioCarta, KEGG, Reactome, WikiPathways, MetaCyc, PID databases as well as 50 hallmarks from MSigDB. In addition, we used immunologic MSigDB collection C7 of 4872 signatures.

Altered gene sets (AGS) from cells and patients

For each sample, a specific AGS was compiled as a list of 25 genes most deviating by expression values from the rest of samples in the respective cohort (either cell lines or patients).

Network visualization

The sub-networks for illustrations were generated at the public NEA resource https://www.evinet.org/ (24).

Least squares models for patients and cells

R packages base, car, nlme were employed, so that the models are presented using R code and functions of these packages.

Model PA-1 (repeated measures ANOVA)

```
gls( model =
Expression ~ Origin + Concentration + Type + Concentration * Type,
correlation = corAR1(
form = ~ Type | Patients,
value = ACF(
        gls(model = Expression ~ Origin + Concentration + Type + Concentration * Type),
        form = ~ Levels | Patients
        )[2,2]
                        ), method="REML");
```

Expression: gene expression;

Concentration: RNA concentration in the samples;

Origin: Oslo or Stockholm site;

Type: baseline or progression.

Model CL-1 (2-way ANOVA with interaction term)

```
aov(Expression ~ Line + Type + Line * Type)
```

with variables

Expression: gene expression;

Line: parental cell line, HCC827 or NCIH1975;

Type: original or resistant.

Model CCLE-1 (3-way ANOVA with one interaction term)

```
anova(lm(Sensitivity ~ Tissue + EGFR + Expression + EGFR * Expression)),
```

with factors

Sensitivity: to one of [ER, GE, OS];

Tissue: tissue or organ of the original tumor;

EGFR: mutation status;

Expression: gene expression.

Model PCA-1 (3-way ANOVA without interaction terms)

```
anova(lm(PC ~ Origin + Concentration + Type)),
```

© Translational Lung Cancer Research. All rights reserved.

Table 1 Clinical parameters

Assessment	Value
Gender (%)	Male: 5/20 (25%), female: 15/20 (75%)
Median age, years (range)	64 (38–79)
Smoking status (%)	Current: 2/20 (10%), ex-smoker: 9/20 (45%), never smoker: 9/20 (45%)
Performance status (%)	0: 7/20 (35%), 1: 12/20 (60%), 2: 1/20 (5%)
Histology (%)	Adenocarcinoma: 20/20 (100%)
Stage at diagnosis (%)	IV: 20/20 (100%)
EGFR mutation subtype (%)	Exon 19: 15/20 (75%), L858R: 4/20 (20%), Exon 18: 1/20 (5%), T790M: 20/20 (100%)
Median PFS osimertinib (months)	10.6

EGFR, epidermal growth factor receptor; PFS, progression-free survival.

with variables

PC: sample-specific principal component value;

Concentration: RNA concentration in the samples;

Origin: Oslo or Stockholm;

Type: baseline or progression.

Ethical statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and the ICH-Guidelines of Good Clinical Practice and according to regulatory requirements. This study received ethical approval by the institutional review board at Karolinska University Hospital (registration No. 2016/944-31/1) and Oslo North Regional Ethics Board (No. 2015/181). Additional approvals by Stockholm Medical Biobank (No. Bbk-01605) were received. All patients provided written informed consent.

Data availability statement

The data generated in this study are available within the article and its supplementary data files. Raw data for this study were generated at the Karolinska Institutet BEA core facility and is available from the corresponding author upon request.

Sample definition and in-laboratory replication

All experiments were conducted using biological replicates. Visualized data reflects either all biological replicates, or representative biological replicates, as stated. If visualized with error bars, each data point represents all biological replicates of a specific analysis group.

Results

Transcript coverage of plasma-derived exosomal RNA

To demonstrate that blood plasma serves as a comprehensive RNA source, we analyzed exosomal RNA from plasma sampled at baseline and progression of disease from twenty EGFR-mutant NSCLC patients receiving osimertinib in a multicenter phase II study (20) (*Table 1*, Table S1) as well as RNA derived from six EGFR-mutant NSCLC cell lines (two parental and four TKI-refractory) (21). First, we compared the level of per chromosome representation of mRNA transcripts detected in our blood plasma and cell line samples to RNA-seq transcriptomics data from traditionally used cancer samples: either *in vitro* Cancer Cell Line Encyclopedia (CCLE) cell cultures or 545 primary, fresh-frozen NSCLC tumors (TCGA) (21,25-27) (*Figure 1*). Using as reference a collection of 386 cancer-related genes, we found that detectable gene expression per chromosome was more variable in the blood plasma samples. However, the overall representation of the mRNA-landscape was found fairly similar between the plasma samples (Clariom D platform), our cell line samples (Clariom D platform), and RNA-seq CCLE and The Cancer Genome Atlas (TCGA) samples, with a maximum of 81% detected genes per sample per chromosome (Table S2).

Next, we explored possible variability of transcriptomics data due to known factors in a principal component analysis (PCA) of all the 40 samples (Figures S1,S2,

© Translational Lung Cancer Research. All rights reserved.

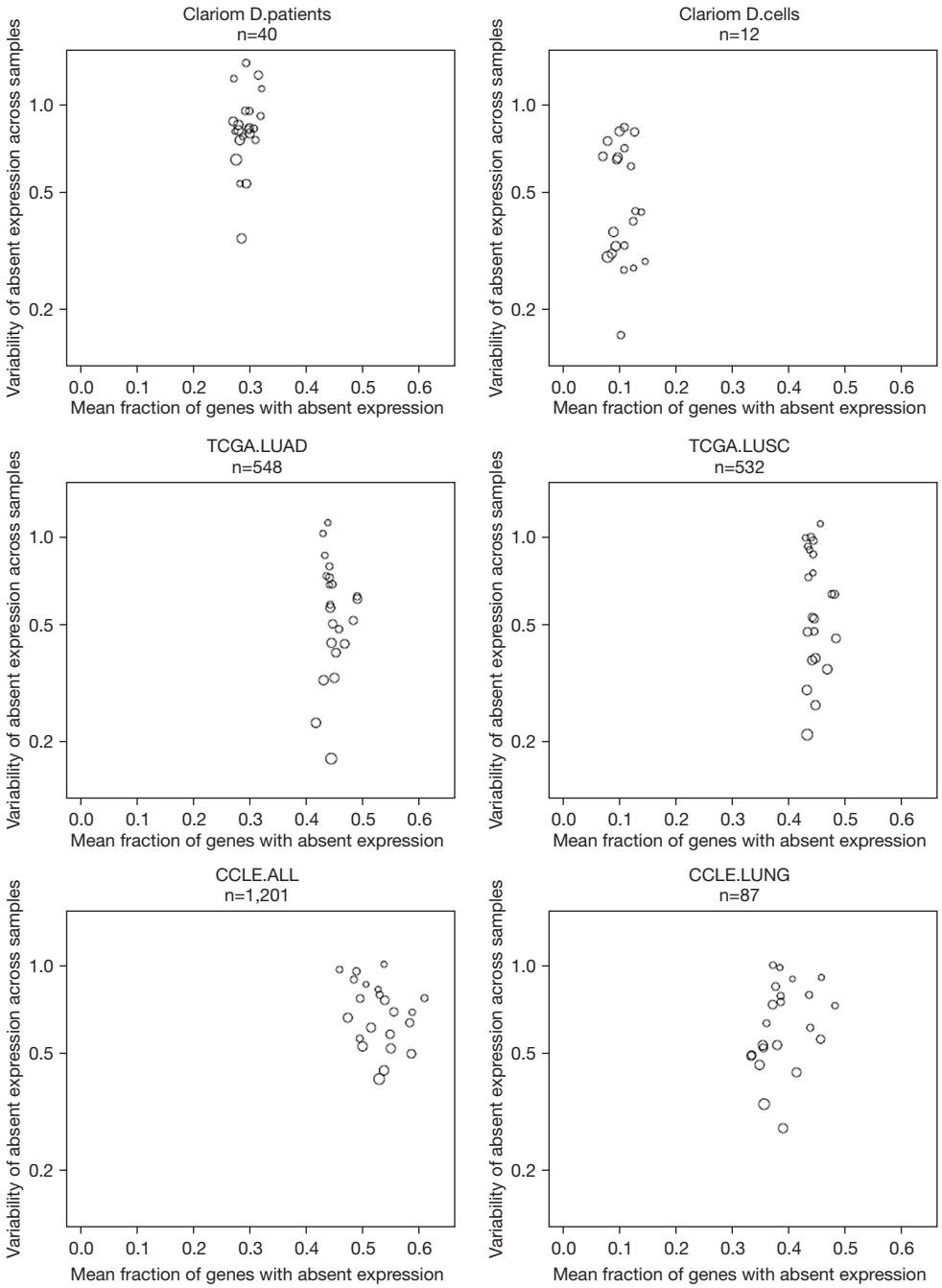

Figure 1 Representation of cancer gene mRNA in transcriptomics datasets. Each circle of the individual panels represents one chromosome. TCGA, The Cancer Genome Atlas; CCLE, cancer cell line encyclopedia; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

Table S3). In order to detect potential influence of sample RNA concentration, site of delivery ("origin", i.e., Stockholm or Oslo) as well as sample type ("baseline *vs.* progression") on specific principal components (PC), we subjected each PC to linear model analysis (model PCA-1, see Methods). While most components were not associated with any changes between baseline and progression, PC32 clearly separated baseline from progression samples

© Translational Lung Cancer Research. All rights reserved.

(Figures S1,S2). This PCA investigation confirmed that influence of the three factors on variability in RNA expression should be accounted for. Therefore, the subsequent analysis of differential expression between baseline and progression included necessary covariates and an interaction term "Concentration * Type" (model PA-1). As an example, we included paired sample information in a repeated measures model on individual genes (model PA-1) and found that most informative genes from PC32 significantly overlapped with genes detected by this model. Differential RNA expression was most pronounced for *PYY3*, *ABCA2*, *MT1L*, *PRODH2*, *HMGB1P19*, *MIR892B* and *OR56B2P* (FDR <0.0001) (Table S4).

Dynamics of plasma-derived exosomal RNA

Based on the conclusions above, we evaluated differential RNA abundance in plasma derived exosomes between baseline and progression samples using a repeated measures model PA-1, which accounted for patients' identities and detected changes due to tumor progression while subtracting influence of total RNA concentration and batch effect of delivery site. We required the interaction term "concentration * progression/baseline" to be insignificant in order to exclude less stable findings. In total, 128 transcripts displayed significant differential expression [log2(fold-change) >0.25, FDR P<0.15, value for interaction term "concentration X progression/baseline" >0.05] (*Figure 2A*, Table S5). Among these, expression of 41 and 87 transcripts decreased and increased toward progression, respectively. Importantly, the most pronounced genes from PC32 (Table S4) were among the top differentially expressed genes. In addition, we detected multiple genes in cell signaling pathway, immune system pathway and transcription, including *MKNK1*, *RASA1*, *RGS18*, *IL17RA*, *ZNF17* and *LIN9* (*Figure 2B-2I*, Table S5).

Pathway enrichment of differentially expressed genes

The differential expression analysis presented above produced a list of genes altered during the treatment. In order to characterize the list at a more general level we subjected the DEG list as an altered gene set (AGS) to network enrichment analysis (NEA) against a collection of 6,529 functional gene sets (FGS). The NEA approach (18) is similar to over-representation analysis of DEG but considers the network context of each gene (*Figure 3*). This overall NEA exposed a number of highly enriched pathways (FDR $<3\times10^{-10}$), including pathways related to ERBB and PI3K signaling, as well as syndecan and glypican pathways. The glypican pathway produced the highest number of AGS to FGS links. Interestingly, we also observed several immune-related pathways, including IFN-gamma and IL-6 signaling pathways (*Figure 4A,4B*, Table S6). The network analysis considers transcripts present in the global interaction network, i.e., nearly all protein coding genes, most miRNAs, but neither long intergenic non-coding RNAs nor pseudogenes. A minor fraction of the pathway genes were identified as DEGs, which emphasizes that enrichment methods not using network analysis would be unlikely to detect these relations.

Differential pathway activation in plasma-derived exosomal RNA

The overall approach above detected pathways that characterized the DEG list as an integral, coherent gene group. Furthermore, we also created individual, patient-specific altered gene sets (AGS) by gathering genes that differed in each given sample from the cohort gene means (*Figure 3*). We compiled 40 sample-specific AGSs and subjected them to NEA. This produced a matrix of 40× 6,529 enrichment values and enabled using NEA scores in the same way as the original mRNA expression values, with the difference that 6,529 NEA profiles were used instead of RNAs. Namely, we detected differential pathway activation (DPA) for 16 out of 6,529 FGSs [FDR<0.25 and P(interaction) >0.05], also requiring that NEA z-score should exceed 3 in at least one of the 40 samples (*Figure 5*, Figure S3, Table S7). One of the 16 differentially activated pathways, gene set GSE35825, displayed biological overlap with FGSs detected in the overall NEA approach (*Figures 4A,5A*, Tables S6,S7). The original publication by Liu *et al.* presented a transcriptomics dataset comparing IFN-alpha versus IFN-gamma stimulated macrophages derived from mouse bone marrow (28). The data was further processed to present the 200 most differentially expressed genes. In our patient data, the sample-specific gene sets were often linked in the global network to the GSE35825-based set of 200 genes and this linkage, quantified as NEA Z-scores, manifested values in the baseline samples systematically higher than the matching progression samples (*Figure 5A*). Interestingly, one gene of the GSE35825-based set [Endothelin 1 (*EDN1*) (29)] was consistently upregulated in the progression samples (*Figure 5B,5C*). In general, it appeared that immunity-related pathways dominated the

© Translational Lung Cancer Research. All rights reserved.

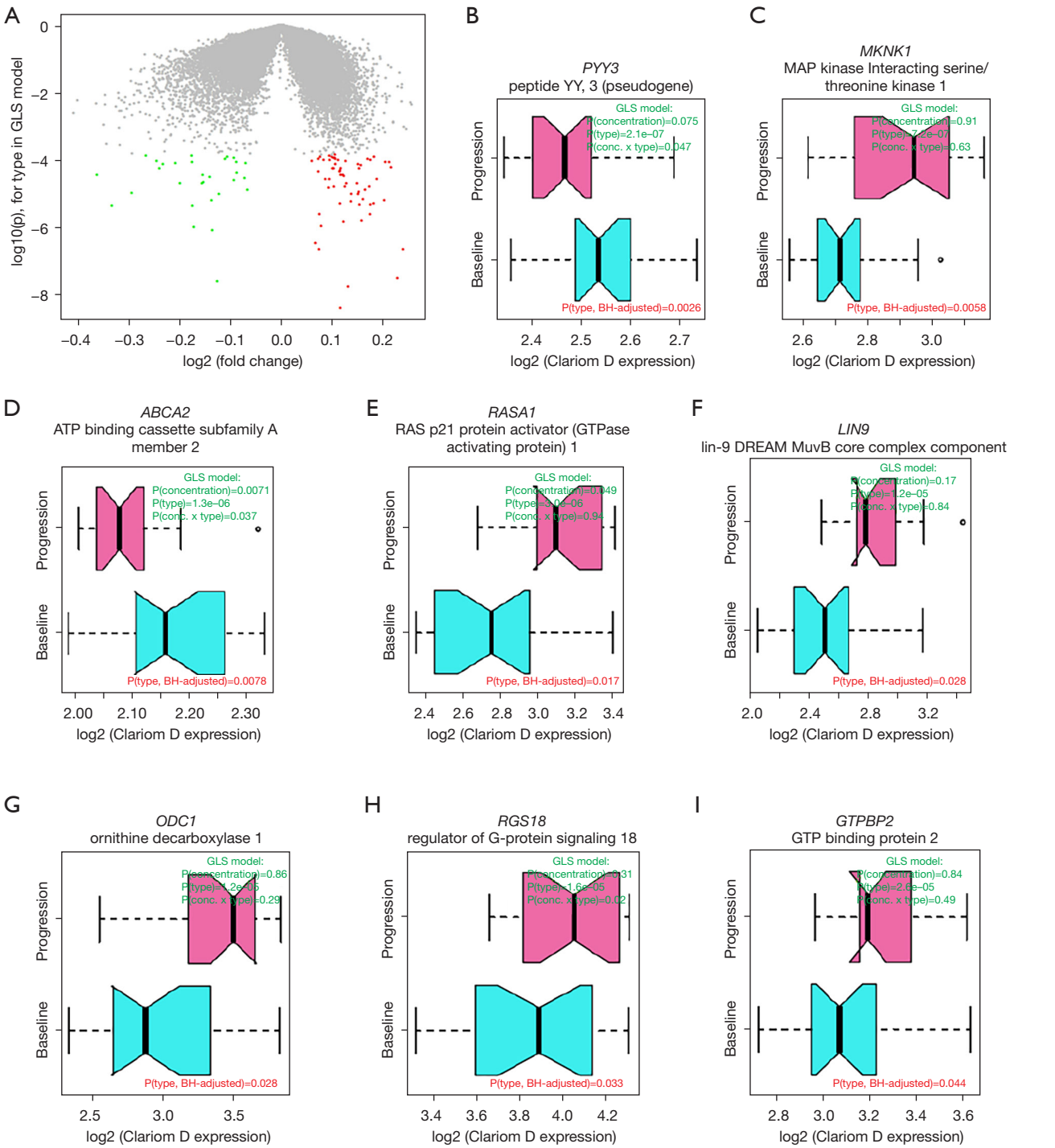

Figure 2 Analysis of differential gene expression of plasma-derived exosomal RNA from NSCLC EGFR T790M patients receiving treatment with osimertinib. (A) Inverted volcano plot of downregulated (green) versus upregulated transcripts (red), significant at Benjamini-Hochberg adjusted P value <0.1, in plasma-derived exosomal RNA from 20 baseline samples versus 20 progression samples. (B-I) Eight representative genes with differential expression post osimertinib treatment. NSCLC, non-small cell lung cancer; EGFR, epidermal growth factor receptor.

© Translational Lung Cancer Research. All rights reserved.

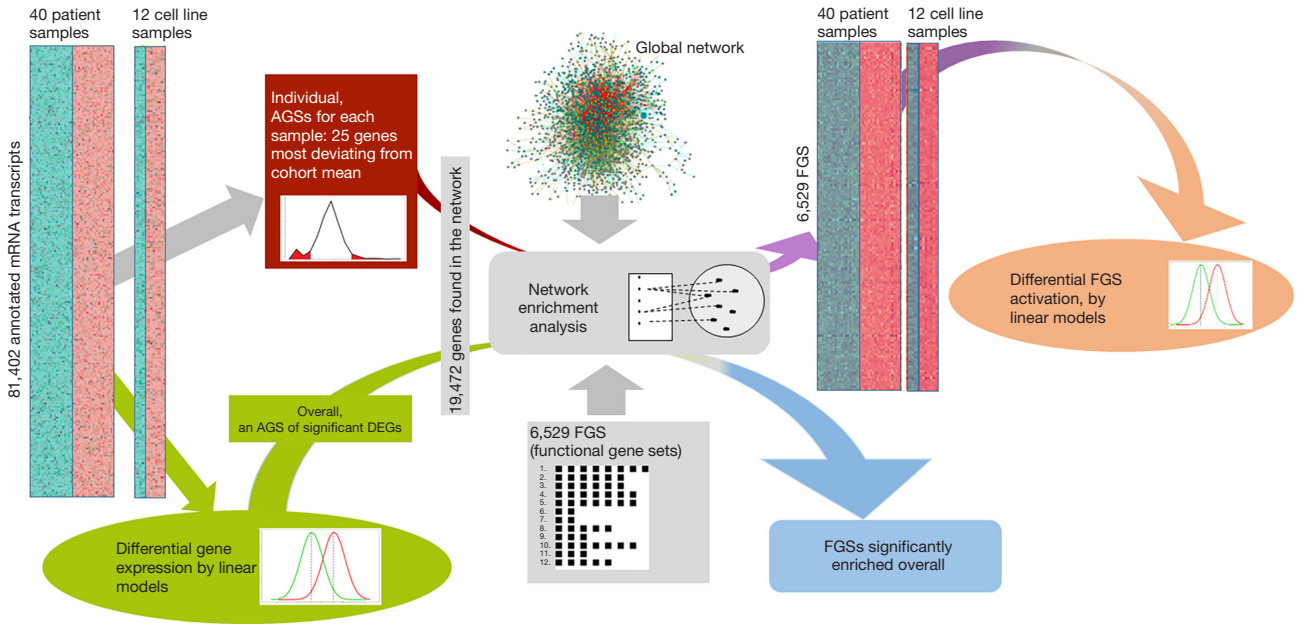

Figure 3 Network analysis. Columns, in each cohort-heatmap, are split into baseline and resistance/progression sample parts (left and right, respectively). Cell line samples represent biological duplicates of six EGFR-mutant NSCLC cell lines (21). NSCLC, non-small cell lung cancer; EGFR, epidermal growth factor receptor; AGS, altered gene sets; FGS, functional gene sets; DEGs, differentially expressed genes.

differences between baseline and progression phenotypes (Table S7).

Reciprocal validation using the three data sources

In order to demonstrate that our findings reflect a potential biological context of acquired resistance upon progression, we used the web resource www.evicor.org (30) in order to match the results obtained in the patient cohort to sensitivity correlates from our TKI-refractory cell line panel and CCLE dataset. We could estimate overlaps between sets of lower p-value correlates between different analyses. This approach, by calculating Fisher's exact statistics of the overlap and controlling error rates via appropriate adjustment for multiple testing, demonstrated that there was a statistically significant match between the findings in all pairwise comparisons and at both gene and pathway levels (Table S8). In total, our cell line panel resulted in 64 enriched pathways related to TKI-resistance versus 33 enriched pathways in the patient cohort (Table S8). Since the experimental setups behind the three data sources were entirely independent, the overlaps indicate a biological and clinical relevance of the patient blood plasma sampling.

Discussion

In this study, we investigated the transcriptome from longitudinally sampled liquid biopsies to assess potential RNA biomarkers with a possible association to the development of resistance to osimertinib in the clinic. While there have been a number of proposed resistance mechanisms to osimertinib, some reports rely solely on analysis of cell-free DNA (31-35), which may not provide the full biological picture of how a tumor can circumvent osimertinib therapy. Our investigation focused on comparing exosomal RNA at treatment baseline to exosomal RNA harvested at disease progression in twenty patients receiving osimertinib as second-line treatment. All patients enrolled in this study had prior treatment with a first-generation EGFR TKI and tested positive for the T790M mutation before starting osimertinib treatment. Therefore, it is possible that the T790M mutation could have arisen as a consequence of prior TKI-treatment and that patients without the T790M mutation would have a distinct RNA landscape in the baseline setting. Moreover, the methodology used in this study (22) is likely to extract RNAs originating from various extracellular vesicles below 200 nm in diameter, and not solely RNAs derived from exosomes. There are so far no studies investigating the

© Translational Lung Cancer Research. All rights reserved.

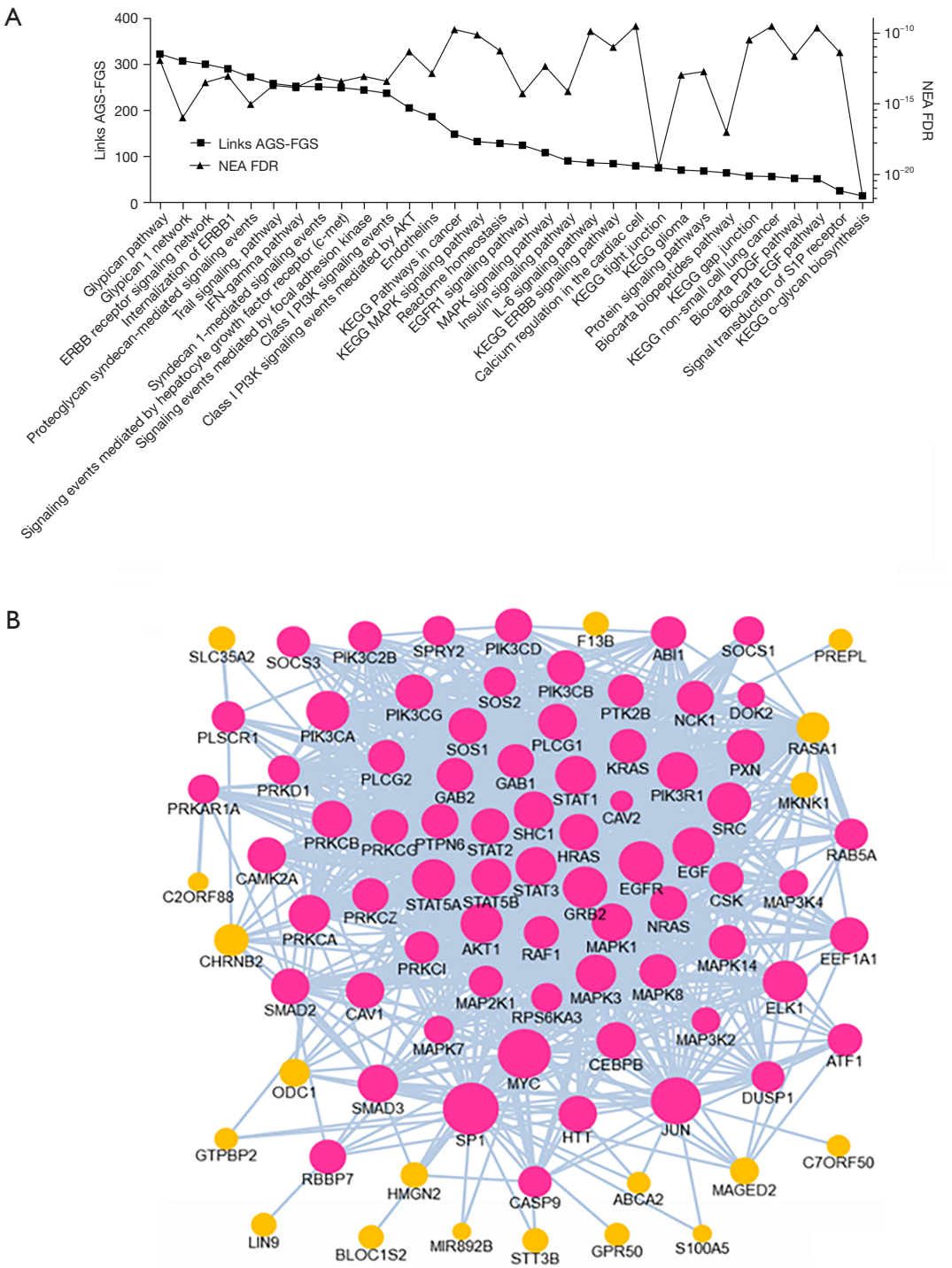

Figure 4 Overall NEA of pathways associated with progression to osimertinib in NSCLC EGFR T790M patients. (A) Graph representation of pathway enrichment, ranked on numbers of individual links between AGS and FGS genes found in the global network. (B) Example of detailed network view of EGFR pathway enrichment versus AGS of most significant DEGs. Yellow: AGS genes; magenta: FGS genes. Genes without links to the opposite set (AGS to FGS and vice versa) are not shown. NEA, network enrichment analysis; NSCLC, non-small cell lung cancer; EGFR, epidermal growth factor receptor; AGS, altered gene sets; FGS, functional gene sets; DEGs, differentially expressed genes; FDR, false discovery rate.

© Translational Lung Cancer Research. All rights reserved.

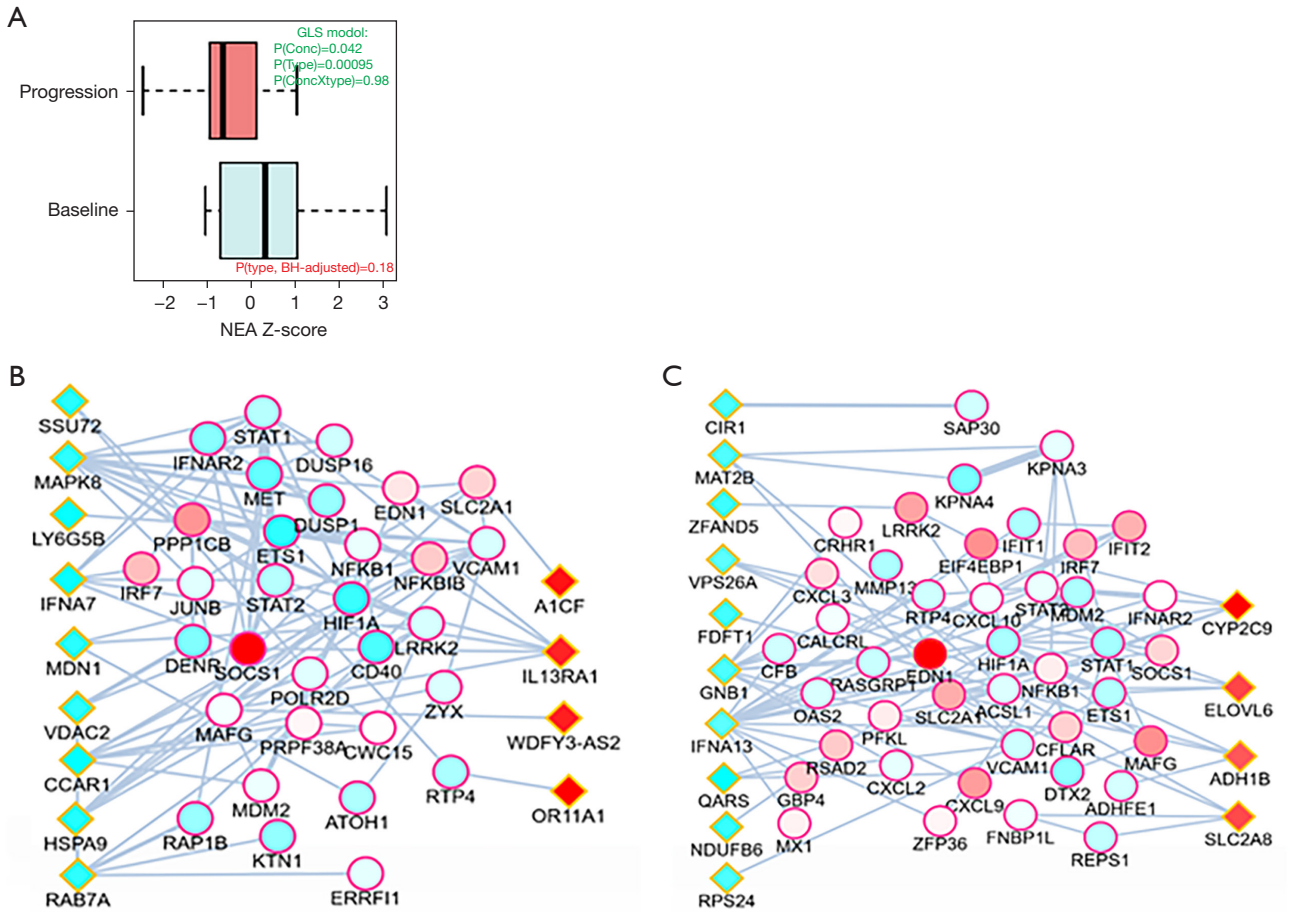

Figure 5 Individual NEA of pathways associated with progression to osimertinib in NSCLC EGFR T790M patients. (A) When subjected to NEA, the 20 baseline AGS showed higher enrichment with respect to GSE35825_IFNA_*VS*._IFNG_STIM_MACROPHAGE_UP compared to progression samples from same patients. (B) Example patterns of network connectivity of baseline AGS from patient 14 (sample 27) versus FGS GSE35825_IFNA_*VS*._IFNG_STIM_MACROPHAGE_UP. Diamonds: AGS genes; circles: FGS genes; shades of blue and red: degree of down- and up-regulation compared to the genes' cohort means, respectively (note that shades of AGS genes are much brighter, since their selection was solely based on differential expression). (C) Same as B, for the progression sample (sample 28). NEA, network enrichment analysis; NSCLC, non-small cell lung cancer; EGFR, epidermal growth factor receptor; AGS, altered gene sets; FGS, functional gene sets.

exosomal transcriptome from plasma, following osimertinib resistance. Analysis of our non-sequencing transcriptomics platform revealed presence of all major RNA categories, which provided a robust base for direct interpretation of results in DEG analysis and for enrichment analyses. The ability to derive biologically sensible results, validated with external experimental and model datasets, provided a proof of potential usability of blood plasma sampling. When comparing the full cohort of baseline versus progression samples, we were intrigued by the relatively low amount of differentially expressed transcripts. We observed significant differential expression of protein coding mRNA transcripts *MKNK1*, *ABCA2*, *PRODH2*, *RASA1*, *IL17RA*, *ZNF17*, *LIN9*, *RGS18*, *APOBEC3D*, *GTPBP2*, *WDR89*, *ODC1*, *ERICH6* and *GSG2*. However, despite the abundant RNA coverage, there was a systematic lack of previously reported aberrations, including *MET*, *HER2* and *PIK3CA*, which might be explained by tumor heterogeneity. This created the incentive of analyzing tumor progression at the pathway level. Interestingly, we observed changes in ERBB, PI3K and ECM (syndecan and glypican) pathways when using the overall NEA approach, while the individual analysis proved

© Translational Lung Cancer Research. All rights reserved.

to be highly informative on presence and involvement of immune cell transcriptomes. Both syndecans and glypicans are cell surface bound heparan sulfate proteoglycans (HSPGs). HSPGs are implicated in regulation of cell proliferation, migration, and differentiation and are therefore considered key players in cancer initiation and progression. However, there is very limited data on the potential role of HSPG in resistance to EGFR TKIs. Nishio *et al.* reported that high concentrations of heparan sulfate in serum were strongly related to poor treatment outcome of EGFR TKIs (36). Heparin-binding epidermal growth factor-like growth factor (HB-EGF) is a ligand for EGFR and has the ability to bind HSPGs, which facilitates EGFR activation. Another study demonstrated that the expression of HB-EGF was clearly increased in lung cancer cell lines with EGFR mutation compared to those without EGFR mutation and implicated HB-EGF as a target in resistance to EGFR TKIs due to EGFR downstream aberrations (37). The role of overexpression of HSPGs in relation to HB-EGF-mediated EGFR activation in TKI resistance remains to be shown. Our study suggests a possible usability of HSPGs as biomarkers in patients with disease progression on osimertinib treatment.

The individual NEA exposed potential roles of immunity-related FGS in the course of progression. This result is in line with a study by Isomoto *et al.* (38), showing that the densities of $CD8^+$ and $FOXP3^+$ lymphocytes as well as the expression of CD73 in tumor cells increased after the development of EGFR TKI resistance, suggestive of possible immunosuppressive effects of regulatory T (Treg) cells and CD73 expression, the latter via induction of adenosine that interacts with the A2A receptor. Notably, most of our findings (either DEGs or differentially activated pathways and gene sets) were significant correlates, i.e., differential values were observed in subsets of samples, which emphasized the complexity of the alterations and the necessity to consider subset and multivariate approaches when developing potential biomarkers.

Profiling liquid biopsies at the RNA level instead of at the DNA level raises some concerns, which need careful consideration. Exosomes are shed into the bloodstream by virtually all cells in the body (39). Therefore, the profiled RNA-landscape will be a mixture of tumor-derived exosomes and exosomes of various sources of non-malignant cells. Although longitudinal profiling will likely reduce the influence from non-malignant cell derived exosomes, the ultimate impact of such influence may look very different from patient to patient, and hence contribute to the observed heterogeneity. On the other hand, the heterogeneity and robustness of mRNA detection in this study was comparable or better than in the public RNA-seq based datasets. In order to truly decipher the impact on, and possibly contribution from, the RNA-landscape in acquired resistance to osimertinib, it may be crucial to extend the analysis to include both liquid biopsies and solid biopsies, as well as to include analysis of DNA and circulating DNA. Such future extension of the analysis, where the RNA-landscape of solid biopsies is compared with the RNA-landscape of liquid biopsies at baseline and progression of disease should determine the ultimate usability of liquid biopsy RNA-profiles in resistance to targeted therapies. Finally, the transcripts and profiles unveiled in this study do not present a causal relationship to osimertinib resistance. Future studies, *in vitro* and *in vivo*, are warranted to validate whether any of the uncovered differentially expressed transcripts play a mechanistic role in circumventing sensitivity to the EGFR TKI osimertinib.

Conclusions

In conclusion, we demonstrate a potential usability of conducting exosomal RNA profiling from plasma to define patients with resistance signatures to the third-generation EGFR TKI osimertinib. Our study highlights the abundance of RNAs in blood plasma, relevance of network-based analysis, and the involvement of multiple RNA species in dictating the transcriptional landscape of osimertinib-refractory NSCLC patients, including mechanisms related to ERBB, ECM and immune-related pathways.

Acknowledgments

We are grateful to the staff at the BEA core facility at Karolinska Institutet for assistance with generation of transcriptomics raw data using the Clariom D platform.
Funding: This study was supported with funding from the Swedish Research Council #2019-01711 (to PH), the Stockholm Cancer Society (to PH, SE), the Swedish Cancer Society (to SE) and the Sjöberg Foundation (to SE).

Footnote

Reporting Checklist: The authors have completed the MDAR

© Translational Lung Cancer Research. All rights reserved.

reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-236/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-236/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-236/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and the ICH-Guidelines of Good Clinical Practice and according to regulatory requirements. This study received ethical approval by the institutional review board at Karolinska University Hospital (registration No. 2016/944-31/1) and Oslo North Regional Ethics Board (No. 2015/181). Additional approvals by Stockholm Medical Biobank (No. Bbk-01605) were received. All patients provided written informed consent.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497-500.
3. Sharma SV, Bell DW, Settleman J, et al. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 2007;7:169-81.
4. Jackman DM, Yeap BY, Sequist LV, et al. Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib. Clin Cancer Res 2006;12:3908-14.
5. Riely GJ, Pao W, Pham D, et al. Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. Clin Cancer Res 2006;12:839-44.
6. Carey KD, Garton AJ, Romero MS, et al. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. Cancer Res 2006;66:8163-71.
7. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med 2005;2:e73.
8. Yun CH, Mengwasser KE, Toms AV, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A 2008;105:2070-5.
9. Cross DA, Ashton SE, Ghiorghiu S, et al. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. Cancer Discov 2014;4:1046-61.
10. Jänne PA, Yang JC, Kim DW, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med 2015;372:1689-99.
11. Yang JC, Ahn MJ, Kim DW, et al. Osimertinib in Pretreated T790M-Positive Advanced Non-Small-Cell Lung Cancer: AURA Study Phase II Extension Component. J Clin Oncol 2017;35:1288-96.
12. Goss G, Tsai CM, Shepherd FA, et al. Osimertinib for pretreated EGFR Thr790Met-positive advanced non-small-cell lung cancer (AURA2): a multicentre, open-label, single-arm, phase 2 study. Lancet Oncol 2016;17:1643-52.
13. Mok TS, Wu Y-L, Ahn M-J, et al. Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. N Engl J Med 2017;376:629-40.
14. Ramalingam SS, Vansteenkiste J, Planchard D, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. N Engl J Med 2020;382:41-50.
15. Papadimitrakopoulou VA, Wu YL, Han JY, et al. Analysis

© Translational Lung Cancer Research. All rights reserved.

of resistance mechanisms to osimertinib in patients with EGFR T790M advanced NSCLC from the AURA3 study. Ann Oncol 2018;29:741.
16. Leonetti A, Sharma S, Minari R, et al. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. Br J Cancer 2019;121:725-37.
17. Mehlman C, Cadranel J, Rousseau-Bussac G, et al. Resistance mechanisms to osimertinib in EGFR-mutated advanced non-small-cell lung cancer: A multicentric retrospective French study. Lung Cancer 2019;137:149-56.
18. Alexeyenko A, Lee W, Pernemalm M, et al. Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. BMC Bioinformatics 2012;13:226.
19. Franco M, Jeggari A, Peuget S, et al. Prediction of response to anti-cancer drugs becomes robust via network integration of molecular data. Sci Rep 2019;9:2379.
20. Eide IJZ, Helland Å, Ekman S, et al. Osimertinib in T790M-positive and -negative patients with EGFR-mutated advanced non-small cell lung cancer (the TREM-study). Lung Cancer 2020;143:27-35.
21. McGowan M, Kleinberg L, Halvorsen AR, et al. NSCLC depend upon YAP expression and nuclear localization after acquiring resistance to EGFR inhibitors. Genes Cancer 2017;8:497-504.
22. Enderle D, Spiel A, Coticchia CM, et al. Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. PLoS One 2015;10:e0136133.
23. Cerami EG, Gross BE, Demir E, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 2011;39:D685-90.
24. Jeggari A, Alekseenko Z, Petrov I, et al. EviNet: a web platform for network enrichment analysis with flexible definition of gene sets. Nucleic Acids Res 2018;46:W163-70.
25. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603-7.
26. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489:519-25.
27. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.
28. Liu SY, Sanchez DJ, Aliyari R, et al. Systematic identification of type I and type II interferon-induced antiviral factors. Proc Natl Acad Sci U S A 2012;109:4239-44.
29. Rosanò L, Spinella F, Bagnato A. Endothelin 1 in cancer: biological implications and therapeutic opportunities. Nat Rev Cancer 2013;13:637-51.
30. Petrov I, Alexeyenko A. EviCor: Interactive Web Platform for Exploration of Molecular Features and Response to Anti-cancer Drugs. J Mol Biol 2022;434:167528.
31. Piotrowska Z, Isozaki H, Lennerz JK, et al. Landscape of Acquired Resistance to Osimertinib in EGFR-Mutant NSCLC and Clinical Validation of Combined EGFR and RET Inhibition with Osimertinib and BLU-667 for Acquired RET Fusion. Cancer Discov 2018;8:1529-39.
32. Shi P, Oh YT, Deng L, et al. Overcoming Acquired Resistance to AZD9291, A Third-Generation EGFR Inhibitor, through Modulation of MEK/ERK-Dependent Bim and Mcl-1 Degradation. Clin Cancer Res 2017;23:6567-79.
33. Yang Z, Yang N, Ou Q, et al. Investigating Novel Resistance Mechanisms to Third-Generation EGFR Tyrosine Kinase Inhibitor Osimertinib in Non-Small Cell Lung Cancer Patients. Clin Cancer Res 2018;24:3097-107.
34. Le X, Puri S, Negrao MV, et al. Landscape of EGFR-Dependent and -Independent Resistance Mechanisms to Osimertinib and Continuation Therapy Beyond Progression in EGFR-Mutant NSCLC. Clin Cancer Res 2018;24:6195-203.
35. Oxnard GR, Hu Y, Mileham KF, et al. Assessment of Resistance Mechanisms and Clinical Implications in Patients With EGFR T790M-Positive Lung Cancer and Acquired Resistance to Osimertinib. JAMA Oncol 2018;4:1527-34.
36. Nishio M, Yamanaka T, Matsumoto K, et al. Serum heparan sulfate concentration is correlated with the failure of epidermal growth factor receptor tyrosine kinase inhibitor treatment in patients with lung adenocarcinoma. J Thorac Oncol 2011;6:1889-94.
37. Yotsumoto F, Fukagawa S, Miyata K, et al. HB-EGF Is a Promising Therapeutic Target for Lung Cancer with Secondary Mutation of EGFRT790M. Anticancer Res 2017;37:3825-31.
38. Isomoto K, Haratani K, Hayashi H, et al. Impact of EGFR-TKI Treatment on the Tumor Immune

© Translational Lung Cancer Research. All rights reserved.

Microenvironment in EGFR Mutation-Positive Non-Small Cell Lung Cancer. Clin Cancer Res 2020;26:2037-46.

39. Cui S, Cheng Z, Qin W, et al. Exosomes as a liquid biopsy for lung cancer. Lung Cancer 2018;116:46-54.

Cite this article as: Alexeyenko A, Brustugun OT, Eide IJZ, Gencheva R, Kosibaty Z, Lai Y, de Petris L, Tsakonas G, Grundberg O, Franzen B, Viktorsson K, Lewensohn R, Hydbring P, Ekman S. Plasma RNA profiling unveils transcriptional signatures associated with resistance to osimertinib in EGFR T790M positive non-small cell lung cancer patients. Transl Lung Cancer Res 2022;11(10):2064-2078. doi: 10.21037/tlcr-22-236

© Translational Lung Cancer Research. All rights reserved.

Supplementary

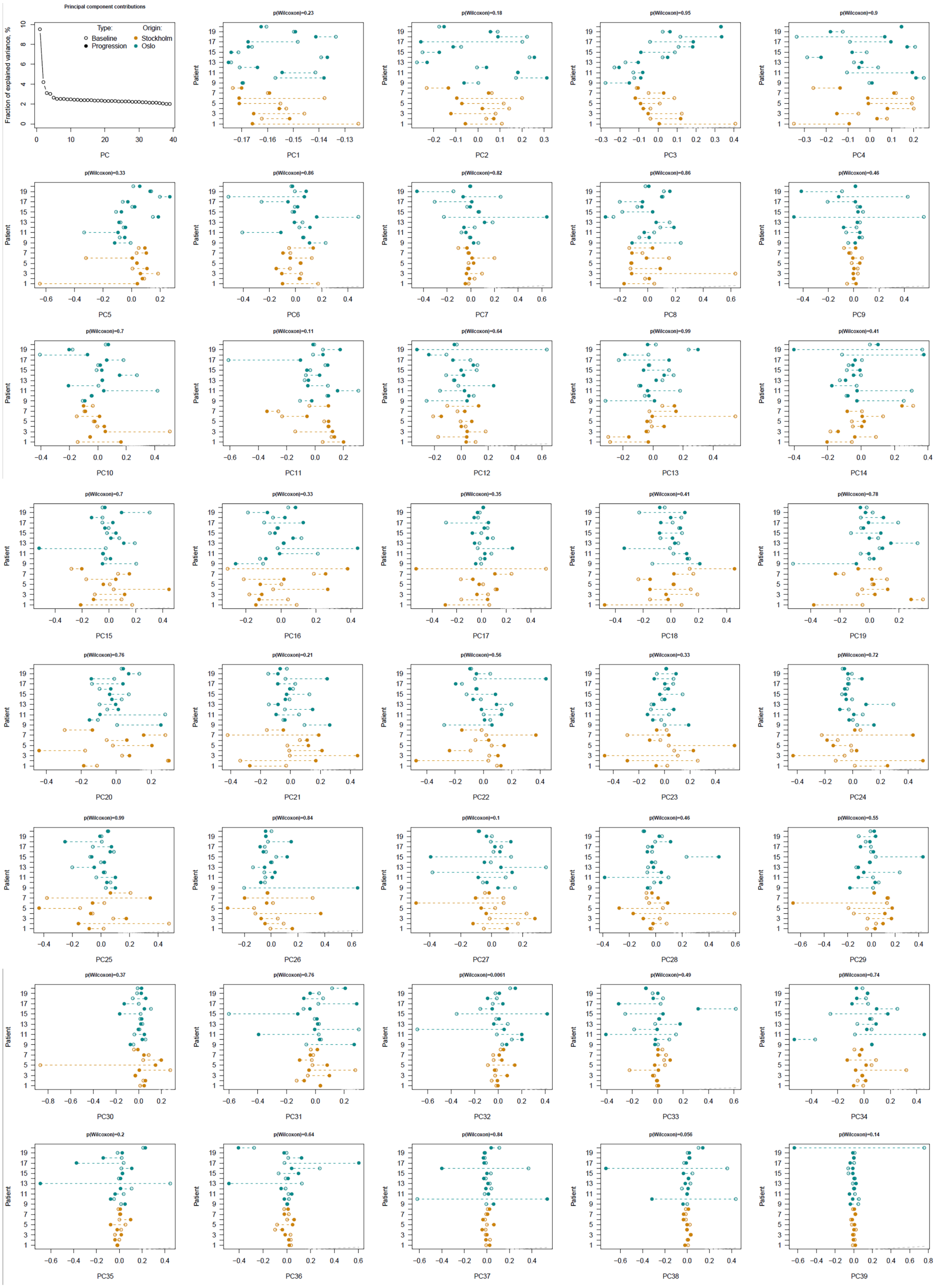

Figure S1 Linear principal component analysis of plasma-derived exosomal RNA across 40 samples from NSCLC EGFR T790M patients receiving treatment with osimertinib.

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

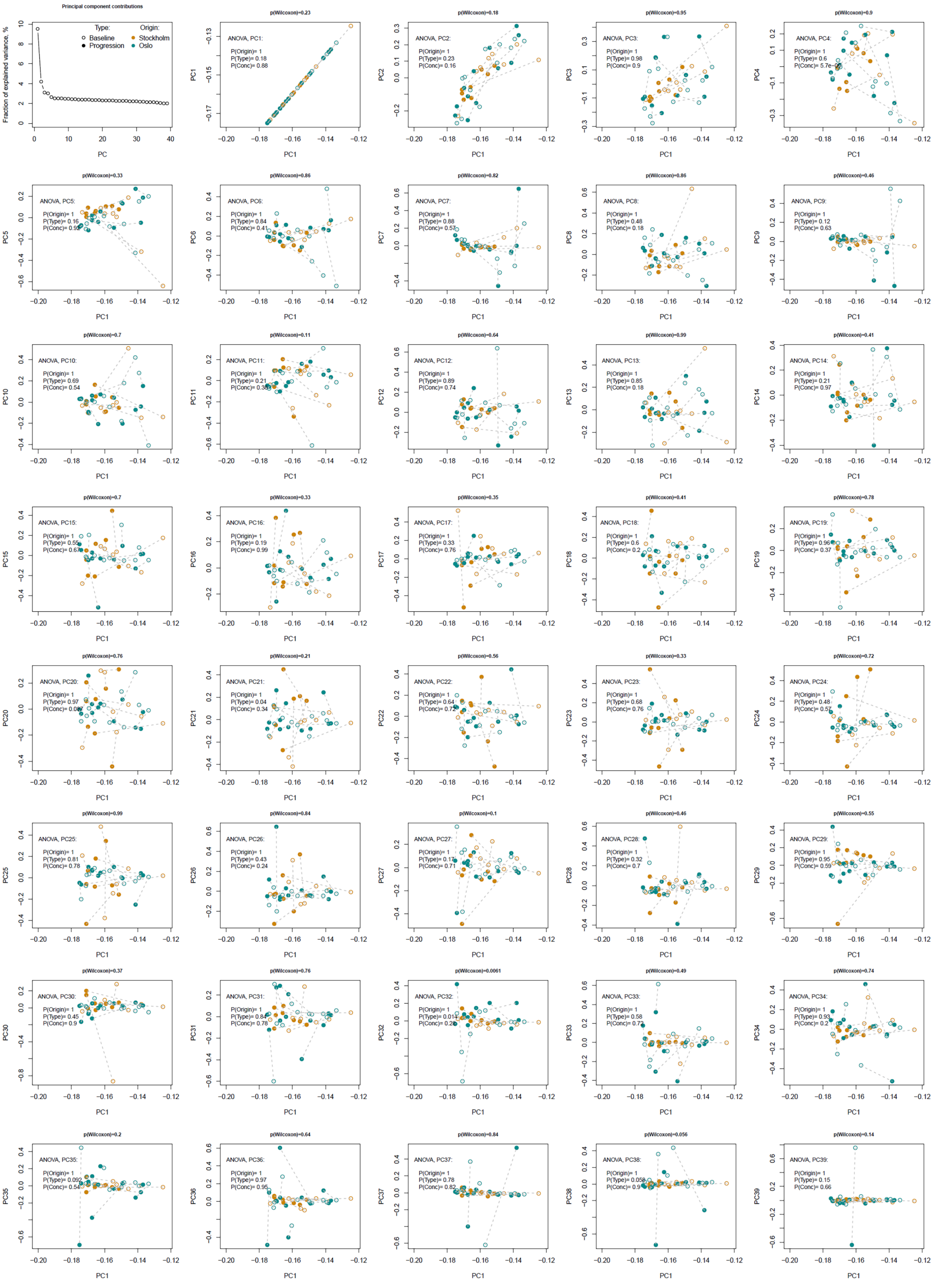

Figure S2 Pair-wise comparison of the 39 principal components.

© Translational Lung Cancer Research. All rights reserved.

https://dx.doi.org/10.21037/tlcr-22-236

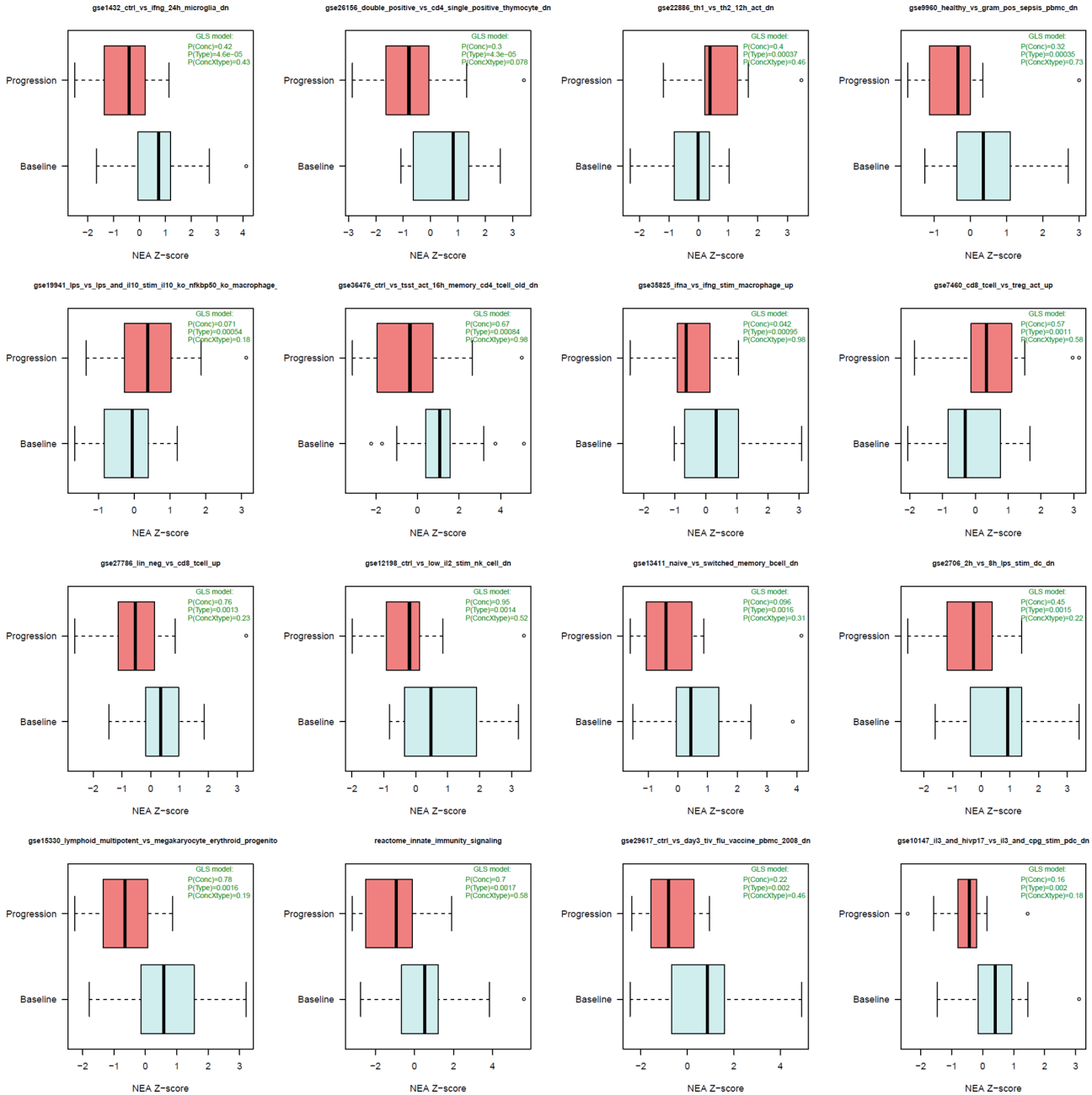

Figure S3 Individual network enrichment analysis (NEA) displaying differentially activated pathways, in addition to GSE35825, with respect to 20 baseline AGS compared to progression samples from the same patient.

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S1 Extended clinical parameters of each individual patient

Site	Progression	T790M-status	Sex	Age	Smoking status	PS	Histol	EGFR mut	PFS months
Helsinki	w16	mut	M	38	Never	1	AC	ex19	3.81
Helsinki	w24	mut	F	62	Never	1	AC	ex19	5.42
Helsinki	w84	mut	M	68	Former	1	AC	ex19	18.73
Linkøping	w32	mut	F	69	Former	2	AC	ex19	7.03
Oslo	w48	mut	F	53	Never	0	AC	ex19	11.3
Oslo	w32	mut	F	73	Current	1	AC	ex18	7.26
Helsinki	w32	mut	F	56	Former	1	AC	ex19	7.2
Linkøping	w24	mut	F	60	Former	1	AC	ex21	5.22
Linkøping	w48	mut	F	75	Former	1	AC	ex21	10.71
Oslo	w32	mut	M	43	Never	1	AC	ex19	7
Oslo	w48	mut	F	79	Former	1	AC	ex21	10.61
Oslo	w60	mut	F	59	Current	1	AC	ex21	13.14
Karolinska	w60	mut	F	75	former	0	AC	ex19	13.73
Karolinska	w32	mut	M	65	never	1	AC	ex19	7.36
Karolinska	w48	mut	F	76	former	0	AC	ex19	10.64
Karolinska	w60	mut	F	66	never	0	AC	ex19	12.88
Karolinska	w32	mut	M	78	never	0	AC	ex19	5.39
Karolinska	w60	mut	F	69	former	0	AC	ex19	14.23
Karolinska	w96	mut	F	61	never	0	AC	ex19	11.24
Karolinska	w16	mut	F	64	never	1	AC	ex19	3.68

EGFR, epidermal growth factor receptor; PFS, progression-free survival; AC, adenocarcinoma.

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S2 Representation of cancer gene mRNA in transcriptomics datasets

Dataset	Source of material	Transcriptomics platform	Mean variability in numbers of genes with low or absent expression across chromosomes	Fraction of genes with low or absent expression across chromosomes, mean ± st.dev.
Clariom_D, patients	Blood plasma	Clariom_D	0.840	0.336±0.03
TCGA.LUAD (lung adenocarcinomas)	Primary tumors, fresh-frozen	RNA-seq	0.567	0.414±0.01
TCGA.LUSC (lung squamous cell sarcomas)	Primary tumors, fresh-frozen	RNA-seq	0.642	0.434±0.02
Clariom_D, cells	Cell culture	Clariom_D	0.521	0.188±0.06
CCLE, all cancer cell lines	Cell culture	RNA-seq	0.698	0.537±0.06
CCLE, cancer cell lines of lung origin	Cell culture	RNA-seq	0.645	0.427±0.05

For each of the human chromosomes, presence of cancer gene transcripts (386 genes found in the KEGG cancer-related pathways; codes KEGG#052*) was evaluated in each sample. Variability was expressed as variance of logit values $\log = \left(\frac{p}{1-p} \right)$, where p was the fraction of cancer genes with low or absent expression. The latter was defined as case of expression in a given sample significantly lower than the cohort mean (at Bonferroni-adjusted P value from *t*-test <0.01).

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S3 Principal component analysis

PC	Eigenvalue's square root	Fraction of variance (%)	P.Type	P.Conc	P.Wilcoxon
1	2.416	9.50	0.181	0.8804	0.23
2	0.467	4.18	0.2269	0.161	0.18
3	0.26	3.12	0.9784	0.9033	0.95
4	0.243	3.02	0.5974	5.74E–06	0.9
5	0.185	2.63	0.1558	0.9167	0.33
6	0.169	2.52	0.8414	0.4125	0.86
7	0.167	2.50	0.8829	0.5654	0.82
8	0.166	2.49	0.4804	0.1768	0.86
9	0.164	2.48	0.1243	0.6255	0.46
10	0.161	2.45	0.6875	0.5357	0.7
11	0.158	2.43	0.2097	0.3815	0.11
12	0.155	2.41	0.8922	0.7445	0.64
13	0.154	2.40	0.8471	0.179	0.99
14	0.152	2.39	0.2138	0.9681	0.41
15	0.152	2.38	0.5484	0.6699	0.7
16	0.15	2.37	0.1854	0.9853	0.33
17	0.148	2.35	0.3339	0.7635	0.35
18	0.148	2.35	0.5988	0.2041	0.41
19	0.144	2.32	0.9628	0.3666	0.78
20	0.143	2.31	0.9695	0.08717	0.76
21	0.142	2.31	0.03978	0.3434	0.21
22	0.14	2.29	0.6438	0.7159	0.56
23	0.139	2.28	0.6827	0.7577	0.33
24	0.138	2.27	0.4772	0.5665	0.72
25	0.137	2.27	0.807	0.7787	0.99
26	0.136	2.26	0.4288	0.2359	0.84
27	0.135	2.25	0.1743	0.7122	0.1
28	0.132	2.22	0.3169	0.7037	0.46
29	0.131	2.21	0.9485	0.5893	0.55
30	0.129	2.19	0.4549	0.9016	0.37
31	0.126	2.17	0.8423	0.7758	0.76
32	0.124	2.16	0.01401	0.2774	0.0061
33	0.123	2.14	0.5784	0.7168	0.49
34	0.122	2.14	0.9263	0.2018	0.74
35	0.121	2.12	0.09156	0.5421	0.2
36	0.118	2.10	0.9653	0.9528	0.64
37	0.111	2.04	0.7752	0.8159	0.84
38	0.109	2.02	0.0585	0.9001	0.056
39	0.105	1.98	0.1471	0.6604	0.14

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S4 Differentially expressed genes in Principal Component 32

Gene	Coefficient in PC32	P.Type	FDR.Type
PYY3	–0.166692315	2.20E–07	9.26E–06
ABCA2	–0.16120866	1.09E–06	1.53E–05
MT1L	0.285908111	8.00E–07	1.53E–05
PRODH2	–0.190786149	1.57E–06	1.65E–05
HMGB1P19	–0.175180873	5.95E–06	4.32E–05
MIR892B	0.1765141	6.17E–06	4.32E–05
OR56B2P	–0.36834929	1.55E–05	9.32E–05
GSG2	–0.322834092	3.73E–05	0.000195949
CHRNB2	–0.16535396	0.000120857	0.000391535
LINC00210	–0.216639704	0.000150933	0.000391535
MAP3K7CL	0.231942374	0.000158478	0.000391535
PGAM1P13	–0.175418078	0.000142233	0.000391535
PIGFP3	0.161248332	0.000119775	0.000391535
PREPL	0.444446945	0.000134088	0.000391535
SLC35A2	0.223094668	0.000128467	0.000391535
STT3B	0.472279445	9.62E–05	0.000391535
ZNF507	0.285862623	0.000131286	0.000391535
SCYL2	0.291051161	0.000174486	0.000407134
CHRNA3	0.167199573	0.000201859	0.000439237
DEFA3	–0.179454824	0.000228262	0.000439237
OR51Q1	–0.280672257	0.000230077	0.000439237
RPL35AP26	0.182099036	0.000218151	0.000439237
PBK	–0.4036554	0.000260394	0.000475502
ATP6V1B2	0.255478305	0.000289924	0.000507367
ATP1B1	0.384966212	0.000335871	0.000542561
SEC22A	0.221573944	0.000324379	0.000542561
ERCC4	0.171045665	0.000417382	0.000649261
FAM228B	0.215143778	0.000444358	0.000666537
GOLGA8B	–0.17285244	0.000471919	0.00068347
CSTF3	0.23327718	0.000578246	0.000792976
GABRG3	0.184682147	0.000585292	0.000792976
MOCS1	–0.177532192	0.000639775	0.000839704
PART1	0.193956082	0.000706772	0.000873072
PCED1A	0.262638443	0.000687793	0.000873072
ARHGAP42	0.265394995	0.000748461	0.000898154
MIR507	–0.233989024	0.000788235	0.000905413
RASSF5	0.185141951	0.000819183	0.000905413
ZNF786	0.228326367	0.000800372	0.000905413
LINC00314	0.289167794	0.000863075	0.000929465
CEP170P1	0.176312692	0.000905245	0.000931036
RNU6-271P	0.196029784	0.000908869	0.000931036
ERGIC2	0.395401712	0.000997413	0.000997413

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S5 Differential gene expression in patients progressing on osimertinib and with an FDR <0.15

Gene_symbol	log2(fold_change).GE_patients	P(ConcXtype).GE_patients	FDR(Type).GE_patients
sworpaw	0.45	0.41	0.00026
luber	−0.36	0.1	0.00048
lawdarbo	0.65	0.15	0.00048
zochorbu	0.71	0.28	0.00048
PYY3	0.29	0.049	0.0023
wawleybo	0.98	0.23	0.0023
siyamu	0.35	0.059	0.0031
MT1L	−0.6	0.22	0.0064
ABCA2	0.25	0.04	0.0069
MKNK1	−0.88	0.58	0.0069
PRODH2	0.64	0.083	0.0083
lokar	0.5	0.66	0.0083
gortor	0.5	0.98	0.011
gertaw	0.58	1	0.011
UPP2-IT1	0.58	0.0015	0.016
sterbybo	0.67	0.23	0.016
terbeyby	0.55	0.23	0.016
IL17RA	−0.66	0.29	0.016
RASA1	−1.72	0.79	0.016
smyku	0.4	0.83	0.016
MIR892B	−0.5	0.37	0.017
choplorby	0.79	0.86	0.017
HMGB1P19	0.44	0.87	0.017
tenura	0.58	0.0067	0.018
CALM2P3	−0.73	0.7	0.022
ZNF17	0.83	0.74	0.022
plawwo	0.41	0.86	0.022
LIN9	−1.29	0.94	0.023
mortyby	0.47	0.15	0.024
tosoru	−0.24	0.043	0.026
gyveebo	0.56	0.84	0.026
OR56B2P	0.39	0.0083	0.029
MIR6818	0.46	0.32	0.029
tinima	0.41	0.75	0.029
slanabo	0.64	0.92	0.029
RGS18	−1.52	0.016	0.034
vasheybu	0.47	0.046	0.034
steywarby	0.48	0.73	0.034
APOBEC3D	−0.94	0.83	0.034
GTPBP2	−0.9	0.44	0.035
TRMT112P2	−0.53	0.3	0.039
floysterby	0.32	0.017	0.04
C18orf65	0.33	0.038	0.04
murera	0.57	0.42	0.04
44451	−0.35	0.07	0.041
WDR89	−0.66	0.2	0.043
spawglu	0.4	0.57	0.044
ODC1	−2.24	0.25	0.046
fustyby	0.54	0.66	0.046
ERICH6	−0.58	0.71	0.046
GSG2	0.46	0.25	0.047
LOC646903	0.32	0.13	0.052
HMGN2P31	−0.45	0.22	0.052
LINC00309	0.37	0.081	0.053
runiyo	0.52	0.067	0.057
geydoy	0.57	0.3	0.06
LINC01101	0.32	0.3	0.06
LINC00358	0.48	0.82	0.06
MICU2	−1.29	0.85	0.06
malor	0.39	0.9	0.06
LOC100131285	0.7	0.32	0.061
gluglybu	0.51	0.31	0.069
H2BFS	−1.11	0.00012	0.074
DDX39BP1	0.56	0.12	0.074
OR8B1P	0.67	0.5	0.074
BLOC1S2	−0.93	0.0088	0.076
EIF4BP1	−0.42	0.34	0.076
nanome	−0.6	0.022	0.078
bluspeeby	0.8	0.43	0.078
LOC606724	−0.41	0.71	0.078
sherbloyby	0.94	0.93	0.078
STT3B	−1.48	0.17	0.082
swawserbu	0.38	0.36	0.082
shermubu	0.51	0.95	0.082
homeobox.49	0.61	1	0.082
showy	0.34	0.23	0.084
RN7SKP294	0.32	0.44	0.084
nergorby	0.57	0.7	0.084
LOC643623	0.31	0.015	0.085
OR51A6P	0.38	0.017	0.085
geegee	0.64	0.3	0.085
HMGN2P32	−0.54	0.37	0.085
RP11-407N8.5	0.38	0.071	0.086
pleyjor	0.31	0.2	0.086
PIGFP3	−0.39	0.099	0.089
CHRNB2	0.38	0.48	0.089
AKAP3	0.74	0.16	0.09
skawspybu	0.42	0.2	0.09
SLC35A2	−0.38	0.33	0.092
PREPL	−1.39	0.24	0.093
ZNF507	−0.78	0.64	0.093
LINC01000	0.5	0.74	0.093
PGAM1P13	0.33	0.77	0.097
MAP3K7CL	−2.03	0.038	0.1
LOC105373185	0.2	0.2	0.1
wawter	0.3	0.42	0.1
C7orf50	−0.31	0.46	0.1
LINC00210	0.39	0.64	0.1
LOC100507443	0.24	0.65	0.1
nawmobu	0.51	0.7	0.1
HMGN2	−0.65	0.0028	0.11
SCYL2	−1.49	0.022	0.11
glarbleeby	0.47	0.081	0.11
RN7SL327P	0.35	0.14	0.11
RN7SL489P	0.35	0.14	0.11
RN7SL853P	0.35	0.14	0.11
OTUD3	0.71	0.16	0.11
S100A5	0.27	0.28	0.11
METTL4	−0.76	0.37	0.11
F13B	−0.52	0.52	0.11
CHRNA3	−0.32	0.66	0.11
rorplarby	0.63	0.73	0.11
TRBV7-7	0.28	0.76	0.11
RPL17P49	−0.44	1	0.11
sarstubu	0.53	0.00042	0.12
OR51Q1	0.64	0.006	0.12
C2orf88	−0.81	0.012	0.12
RPL35AP26	−0.3	0.023	0.12
yosoyu	0.51	0.088	0.12
NPC1	−0.58	0.31	0.12
DEFA3	1.09	0.37	0.12
GPR50	0.41	0.41	0.12
MAGED2	−0.69	0.59	0.12
LOC257396	0.17	0.69	0.12
spodu	0.44	0.78	0.12
nawblabu	0.33	0.79	0.12
LOC730100	0.25	0.83	0.12
shersoby	0.43	0.96	0.12
ELOVL7	−1.79	0.00096	0.13
KIAA1383	0.46	0.019	0.13
wordee	0.51	0.11	0.13
kleylaw	0.61	0.16	0.13
MIR4524B	0.7	0.26	0.13
OR4C13	−0.28	0.31	0.13
fawwee	0.44	0.53	0.13
timemi	−0.51	0.57	0.13
varsperbu	0.48	0.61	0.13
PBK	0.46	0.62	0.13
yanaro	0.39	0.69	0.13
skoplorbu	0.5	0.72	0.13
ATP6V1B2	−1.52	0.91	0.13
dorkluby	0.48	0.017	0.14
DDX11L2	−1.39	0.041	0.14
DDX11L10	−1.49	0.06	0.14
toybo	0.63	0.11	0.14
skugybu	−0.23	0.12	0.14
tukly	−0.48	0.14	0.14
toykoy	0.37	0.14	0.14
ATP1B1	−1.01	0.16	0.14
SEC22A	−0.58	0.22	0.14
NUPR2	−0.26	0.3	0.14
fawchy	−0.27	0.41	0.14
korboybo	0.4	0.69	0.14
steymoybo	0.58	0.69	0.14
CCT8L1P	−0.21	0.79	0.14
slorfer	0.44	0.86	0.14
MLXIP	−0.36	0.86	0.14
kuchoyby	0.8	0.91	0.14
RN7SL70P	−0.34	0.98	0.14
spudu	0.35	0.98	0.14

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S6 Significant pathways in the overall NEA approach

FGS (pathway)	No.of.links.AGS-FGS	NEA.z-score	NEA.P-value	NEA.FDR
glypican_pathway	322	7.681	1.60E-14	1.10E-12
glypican_1_network	307	8.858	8.20E-19	9.60E-17
erbb_receptor_signaling_network	300	8.166	3.20E-16	2.90E-14
internalization_of_erbb1	290	8.028	9.90E-16	8.40E-14
proteoglycan_syndecan-mediated_signaling_events	272	8.601	7.90E-18	8.40E-16
trail_signaling_pathway	258	8.229	1.90E-16	1.80E-14
all623_ding2008	256	12.578	2.80E-36	1.30E-33
ifn-gamma_pathway	252	8.269	1.30E-16	1.30E-14
syndecan-1-mediated_signaling_events	251	8.051	8.20E-16	7.10E-14
signaling_events_mediated_by_hepatocyte_growth_factor_receptor_(c-met)	249	8.14	3.90E-16	3.50E-14
signaling_events_mediated_by_focal_adhesion_kinase	244	8.036	9.30E-16	8.00E-14
class_i_pi3k_signaling_events	237	8.138	4.00E-16	3.60E-14
class_i_pi3k_signaling_events_mediated_by_akt	205	7.496	6.60E-14	4.40E-12
endothelins	186	7.968	1.60E-15	1.30E-13
kegg_05200_pathways_in_cancer	148	6.982	2.90E-12	1.60E-10
kegg_04010_mapk_signaling_pathway	132	7.111	1.20E-12	6.80E-11
reactome_hemostasis	128	7.476	7.60E-14	5.10E-12
hs_egfr1_signaling_pathway_wp437_35716.txt	124	8.39	4.90E-17	4.80E-15
hs_mapk_signaling_pathway_wp382_38878.txt	108	7.82	5.30E-15	4.10E-13
hs_insulin_signaling_wp481_38887.txt	90	8.345	7.10E-17	6.90E-15
hs_il-6_signaling_pathway_wp364_35645.txt	86	7.023	2.20E-12	1.20E-10
kegg_04012_erbb_signaling_pathway	84	7.397	1.40E-13	9.10E-12
hs_calcium_regulation_in_the_cardiac_cell_wp536_38956.txt	79	6.9	5.20E-12	2.80E-10
kegg_04530_tight_junction	75	9.733	2.20E-22	3.60E-20
kegg_05214_glioma	70	8.006	1.20E-15	1.00E-13
hs_g_protein_signaling_pathways_wp35_35311.txt	68	7.934	2.10E-15	1.70E-13
biocarta_biopeptides_pathway	64	9.129	6.90E-20	9.00E-18
kegg_04540_gap_junction	57	7.227	5.00E-13	3.00E-11
kegg_05223_non-small_cell_lung_cancer	56	6.899	5.20E-12	2.80E-10
biocarta_pdgf_pathway	52	7.601	2.90E-14	2.10E-12
biocarta_egf_pathway	51	6.944	3.80E-12	2.10E-10
hs_signal_transduction_of_s1p_receptor_wp26_38752.txt	25	7.513	5.80E-14	3.90E-12
kegg_00512_o-glycan_biosynthesis	14	10.232	1.40E-24	2.80E-22

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236

Table S7 Significant pathways in the individual NEA approach

Gene_set	log2(fold_change)	p(ConcXtype)	FDR(Type)
gse26156_double_positive_*vs.*_cd4_single_positive_thymocyte_dn	1.26	0.078	0.037
gse1432_ctrl_*vs.*_ifng_24h_microglia_dn	1.21	0.43	0.037
gse9960_healthy_*vs.*_gram_pos_sepsis_pbmc_dn	0.84	0.73	0.13
gse22886_th1_*vs.*_th2_12h_act_dn	-0.89	0.46	0.13
gse19941_lps_*vs.*_lps_and_il10_stim_il10_ko_nfkbp50_ko_macrophage_up	-0.61	0.18	0.17
gse36476_ctrl_*vs.*_tsst_act_16h_memory_cd4_tcell_old_dn	1.49	0.98	0.18
gse35825_ifna_*vs.*_ifng_stim_macrophage_up	0.94	0.98	0.18
gse7460_cd8_tcell_*vs.*_treg_act_up	-0.62	0.58	0.19
gse27786_lin_neg_*vs.*_cd8_tcell_up	0.82	0.23	0.21
gse12198_ctrl_*vs.*_low_il2_stim_nk_cell_dn	1.07	0.52	0.21
gse2706_2h_*vs.*_8h_lps_stim_dc_dn	1.04	0.22	0.21
gse13411_naive_*vs.*_switched_memory_bcell_dn	0.93	0.31	0.21
gse15330_lymphoid_multipotent_*vs.*_megakaryocyte_erythroid_progenitor_up	1.23	0.19	0.21
reactome_innate_immunity_signaling	1.56	0.58	0.22
gse10147_il3_and_hivp17_*vs.*_il3_and_cpg_stim_pdc_dn	0.91	0.18	0.23
gse29617_ctrl_*vs.*_day3_tiv_flu_vaccine_pbmc_2008_dn	1.3	0.46	0.23

Table S8 Cross-validation of significant findings

	Patient cohort	In-house cell panel	CCLE
Patient cohort	GE, LM PA-1: 128 DE RNAs NEA(individual), LM PA-1: 16 DA FGS NEA(overall): 33 enriched pathways	NEA(individual): P(H-B) = 0.06 NEA(overall): P(H-B) = 0.011	NEA(individual): P(H-B) = 0.022
In-house cell panel	GE: p(H-B) = 0.0012	GE, LM CL-1: 16 DE RNAs NEA(individual), LM CL-1: 0 DA FGSs NEA(overall): 64 enriched pathways	NEA(individual): P(H-B)=0.0023
CCLE	GE: P(H-B) = 0.0002	GE: P(H-B) = 1.4e–25	GE, LM CCLE-1: 125 DE genes NEA(individual), LM CCLE-1: 0 DA FGS

GE, gene/RNA level analysis; NEA, pathway level analysis; LM, linear model; DE, differentially expressed; DA, differentially activated; P(H-B), P value from Fisher's exact text tried at a series of P value cutoffs of increasing strength, adjusted for multiple testing by Holm-Bonferroni method.

© Translational Lung Cancer Research. All rights reserved.
https://dx.doi.org/10.21037/tlcr-22-236