Sample size calculation for a NanoString GeoMx spatial transcriptomics experiment to study predictors of fibrosis progression in non-alcoholic fatty liver disease

Maria Ryaboshapkina[1+], Vian Azzu[2]

[1] Translational Science and Experimental Medicine, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

[2] Translational Science and Experimental Medicine, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK.

[+]Corresponding author: maria.ryaboshapkina@astrazeneca.com , +46725070342

Abstract

Sample size calculation for spatial transcriptomics is a novel and understudied research topic. Prior publications focused on powering spatial transcriptomics studies to detect specific cell populations or spatially variable expression patterns on tissue slides. However, power calculations for translational or clinical studies often relate to the difference between patient groups, and this is poorly described in the literature. Here, we present a stepwise process for sample size calculation to identify predictors of fibrosis progression in non-alcoholic fatty liver disease as a case study. We illustrate how to infer study hypothesis from prior bulk RNA-sequencing data, gather input requirements and perform a simulation study to estimate required sample size to evaluate gene expression differences between patients with stable fibrosis and fibrosis progressors with NanoString GeoMx Whole Transcriptome Atlas assay.

Key points

Spatial transcriptomics is predominantly used in unpowered exploratory research.

We demonstrate that spatial transcriptomics studies with NanoString GeoMx Whole Transcriptome Atlas can be powered to address differences between patient groups and illustrate the sample size calculation using fibrosis progression in non-alcoholic fatty liver disease as an example.

We note that established software implementing negative binomial mixed effect models lme4 and GLMMadaptive yield more similar fold change estimates to simulated data than the recently published software GeoDiff.

Keywords

Spatial transcriptomics, NanoString GeoMx Whole Transcriptome Atlas, sample size calculation, non-alcoholic fatty liver disease

Introduction

Spatial transcriptomics offers significant advantages over bulk or single cell transcriptomics. Bulk transcriptomics is unable to reveal tissue-wide patterning and disease-promoting cell niches. Single cell omics typically requires fresh tissue, whose dissociation can lead to significant cell type biases. Both these factors can be overcome with spatial transcriptomics, which additionally offers the ability to use widely available formalin-fixed paraffin-embedded (FFPE) tissue.

Two spatial RNA sequencing technologies are commercially available: Visium spatial gene expression from 10X Genomics[1] and GeoMx Digital Spatial Profiler from NanoString [2, 3].

The Visium platform uses functionalized slides with printed barcoded oligo capture probes onto which RNA from permeabilised tissue can bind. This can subsequently be sequenced using a polyA approach. Visium is limited by pre-determined sequenceable spot sizes of 55μm with 100μm spacing between spot centres.

The GeoMX platform extends more flexibility in enabling the operator to select regions of interest/illumination (ROI) from single cell analysis to whole regions based on histology or protein probes. This platform links complementary sequence probes to a unique barcode through a UV-cleavable linker. Hybridised sequence probes on the tissue of interest can then be released from tissue sections through the application of UV only to selected ROIs. The GeoMX platform allows addition of custom targets of interest such as transcript variants. GeoMx Human Whole Transcriptome Atlas enables targeted sequencing of 18,000 genes [3]. While both Visium and NanoString GeoMx are compatible with FFPE tissue, sequencing efficiency with Visium is reduced in FFPE compared to fresh frozen tissue while no such limitation is reported for NanoString GeoMx [4].

Sample size calculation is a crucial step in planning spatial transcriptomics studies. Yet, limited methodological research is available on this topic, especially in clinical specimens. Previous studies reporting sample size calculation for spatial transcriptomics focused on detection or co-localization of specific cell populations on the slide or identification of spatially variable features [5-7]. Specifically,

3

despite the widespread use of the technology in unpowered exploratory research (https://nanostring.com/resources/publications/), only a handful of studies with NanoString GeoMx report dedicated sample size calculation based on pilot experiments but do not outline the requirements and statistical considerations in detail [8, 9].

Here, we illustrate how to gather required inputs (formulate study hypothesis and estimate effect size based on historic data) and perform sample size estimation for comparison between patient groups with NanoSting GeoMx Whole Transcriptome Atlas using fibrosis progression in non-alcoholic fatty liver disease as a case study.

Non-alcoholic fatty liver disease (NAFLD) affects up to a third of the population [10] and covers a spectrum of pathology from simple steatosis (NAFL) to steatohepatitis (NASH) and advanced fibrosis [11]. Progressed NAFLD places patients at risk of liver cancer and is strongly associated with the metabolic syndrome (MetS, comprising obesity, insulin resistance, type 2 diabetes mellitus, hypertension) and coronary heart disease [12].

NAFLD has been shown to progress to fibrosis both in patients with NAFL and NASH [13, 14], with fibrosing disease being the leading predictor of poor outcomes/survival [15, 16]. To date, no studies have been able to show clinical, biochemical or histological predictors of fibrosis progression in cohorts of with serial liver sampling of the same individual [13, 14, 16-19]. A combination of follow-up biopsy analysis and cross-sectional studies indicate that weight gain [13], metabolic co-morbidities such as diabetes [13, 14, 17, 19]and at-risk genotypes, for example in PNPLA3, TM6SF2, MBOAT7 [20, 21] place patients at higher risk of NAFLD fibrosis progression. Non-genetic molecular mechanisms placing the patient at risk of fibrosis progression are particularly poorly elucidated and limited to hepatic gene expression-based or serum protein-based risk signatures also found in HCV fibrosis progression [22]. Thus, further research is required to understand predictors fibrosis progression in NAFLD and, ultimately, adjust clinical management of higher risk patients.

Methods

Initial hypothesis generation

First, we performed analysis to identify which spatial expression patterns could serve as the primary and secondary endpoints.

We explored bulk liver RNA-sequencing cohort from Fujiwara *et al.* (GSE193066 [23]). Fibrosis progressors and fibrosis regressors had at least +1 and -1 fibrosis stage in their circa 2 year follow-up biopsies. The cohort included 28 individuals with stable fibrosis stage, 15 fibrosis regressors and 15 fibrosis progressors. We sought to identify an association between a higher-level gene expression pattern and some element of liver microanatomy or histology. Therefore, we did not make any attempt to control false discovery rate at the level of individual genes or pathways in this analysis. We also used the published DESeq2 rlog-transformed data set as-is without re-processing. However, we did check that results obtained with linear regression on rlog transformed data and on raw counts with DESeq2 [24] were overall consistent, especially with respect to log2 fold change estimates, with this sample size in another cohort GSE135251 [25] where both types of data were readily available (**Figure S1**). Candidate genes were identified with linear regression on rlog data adjusted for baseline fibrosis stage in GSE193066. The genes were required to have opposite direction of log2 fold change in fibrosis progressors and fibrosis regressors compared to stable fibrosis individuals and raw $p < 0.05$ in both comparisons. The candidate genes were semi-manually grouped into broad marker sets using AmiGO 2 [26] and PANTHER [27] based on GO [28, 29] and Reactome [30] pathway definitions and hierarchical relationships between the ontology terms. We calculated a summary score for each of these broad candidate marker sets with Kuppe's method [31]. We verified that the expression of each candidate marker set as a whole, differed between fibrosis progressors, stable fibrosis and fibrosis regressors using linear regression adjusted for both baseline fibrosis stage and baseline NAS score.

The candidate marker sets were derived from bulk liver. As the next step, we traced potential sources of these expression signals within the liver. We calculated summary scores for liver cell types based on markers from GSE136103 [32], hepatocyte zones in NASH liver [33] and for extracellular matrix (ECM, "core matrisome" markers from [31] in the Fujiwara *et al.* cohort (GSE193066 [23]) and in

published Visium liver spatial transcriptomics samples [34-36]. Visium data was SCT-transformed prior to analysis. We assumed "guilt-by-association": if summary scores correlated both in bulk and in spatial transcriptomics samples (Spearman method), then the expression signals co-localized.

Simulation study

Non-tumour liver samples adjacent to hepatocellular carcinoma [34] had well-defined fibrosis areas and hepatocyte fraction and satisfied the 100 reads/μm^2 recommendation by NanoString. We used these five samples for the simulation study.

We defined fibrotic niche as Visium spots that were in the 85 percentile or above for ECM score and at or below the 15 percentile of the hepatocyte score. Similarly, hepatocyte spots were at or above the 20 percentile for hepatocyte score and at or below the 80 percentile of the ECM score. The remaining spots were enriched in other cell types (e.g., cholangiocytes) or had low cell density. This resulted in 2.6-7.1% fibrotic niche, which was within the NAFLD fibrosis range [37], and 63-68% hepatocyte area, which corresponded to the lower range of literature estimates for the hepatocyte fraction by liver volume [38-40]. In total, we obtained 819 fibrotic niche spots and 12,074 hepatocyte spots across the 5 samples. To verify that our approach to tissue segmentation was reasonable, we conducted differential expression analysis between "bulk" tissue constructed by summing up counts from fibrotic niche and hepatocyte spots per patient. As expected, fibrotic niche was enriched in markers of diverse non-parenchymal cell types (mesenchymal, immune etc) and hepatocyte fraction was enriched in hepatocyte markers (**Data S1**).

We translated the expression changes in summary scores from bulk tissue into expression changes within fibrotic niche and hepatocyte fraction. We found two representative markers PON1 and FLNA that had the most similar spatial expression pattern to the two summary scores identified in initial analysis. All subsequent calculations were based on these two genes.

We spiked-in increasingly large fractions of read counts to PON1 in hepatocytes and FLNA in the fibrotic niche. For example, spiked-in fraction r0.1 for FLNA corresponded to FLNA read counts in each fibrotic niche spot in sample i + 0.1*median(FLNA read counts across fibrotic niche spots in

6

sample i), where i =1…5. We summed up read counts across all spots per sample to obtain "bulk" tissue. We calculated log2 fold change between spiked-in "bulk" and unmodified "bulk" tissue with DESeq2 using the patient as a fixed effect covariate and constructed calibration curves. We found the point on each calibration curve where fold change in "bulk" tissue matched log2 fold change observed in the initial hypothesis-generating analysis in the Fujiwara *et al.* cohort [23]. This determined the magnitude of spike-in that we used in the simulation study.

We took the unmodified 12,074 hepatocyte spots and 819 fibrotic niche spots across the five Visium samples as our spot "population" in "stable fibrosis" individuals. The same spots but spiked-in with fractional read counts r0.43 for PON1 and r2.9 for FLNA, respectively, were our spot "population" in "fibrosis regressors" and "fibrosis progressors", respectively.

Initial Visium spot diameter of 55μm contains 8 to 20 cells [34] and is not recommended by NanoString because it does not support reliable differential expression ("The GeoMx® Human Whole Transcriptome Atlas for the Digital Spatial Profiler: Design, Performance, and Experimental Guidelines" document from www.nanostring.com/GeoMxDSP). Fibrotic niche was the smaller tissue fraction in our experimental plan and it determined the range of diameters that we evaluated. The width of fibrotic niche in the refence data rarely exceeded 2-3 Visium spots (55μm*2 + 45 μm distance between spots = 155μm, 55μm*3 + 2*45μm = 255μm). Reference estimates of fibrotic niche size reported in NAFLD were in the <200μm range [37, 41, 42]. Based on this prior data, we considered 165μm as maximum spot diameter. The amount of input material (cells) is determined by tissue area that is subjected to sequencing. The relationship between spot diameter and area is quadratic, so we aggregated 2 Visium spots to mimic 80μm diameter, 4 Visium spots to mimic 110μm, and 9 Visium spots to mimic 165μm.

In NanoString terminology, a spot is called "region of illumination", or ROI, and corresponds to the small piece of tissue from which barcodes are harvested for sequencing. We use the term "ROI" to describe the synthetic larger-diameter NanoString data and the term "spot" to describe original 55μm Visium spots.

We drew random subsamples from our unmodified and spiked-in Visium spot "populations" and summed up counts from neighbouring spots to mimic NanoString ROIs with larger diameter. We made 100 synthetic data sets for each condition (endpoint x N patients x N ROIs x ROI diameter). In the original Visium data, correlations between global gene expression profiles in spots within one patient were not stronger on average than between spots from different patients. So, the new synthetic larger ROIs were randomly assigned to new "patient" or a new ROI from an existing "patient". The simulated data sets were restricted to genes measured on the Nanostring GeoMx Human Whole Transcriptome Atlas assay (GPL32201). Genes that were part of the assay and had median expression <0.05 TPM in all GTEx tissues (https://gtexportal.org/home/ [43]) were used as negative control "background probes" for the GeoDiff method [44].

Since spatial transcriptomics is a new technology and the NanoString software GeoDiff [44] has not been extensively validated yet, we wanted to increase confidence in the results and applied three methods to test differences in expression levels of FLNA and PON1 between spiked-in and unmodified data sets. Conceptually, all three methods relied on negative binomial mixed effect regression models with background or effective library size treated as a threshold or an offset. Therefore, expression of a gene of interest was corrected for background. In all cases, patient group was treated as the fixed effect (e.g. "fibrosis progressor" vs "stable fibrosis") and patient as the random intercept. Effective library size was calculated with edgeR [45].

Method 1, GeoDiff [44]:

GeoDiff::fitNBthmDE with gene count ~ group + (1|patient) after following the processing steps including the background estimation as described in the package vignette (https://bioconductor.org/packages/release/bioc/vignettes/GeoDiff/inst/doc/Workflow_WTA_kidney.html).

Method2, glmer.nb [46]:

lme4::glmer.nb with gene count ~ offset(natural logarithm of effective library size) + group + (1|patient)

8

Method 3, GLMMadaptive LRT [47, 48]:

Likelihood ratio test (LRT) between full model assuming that there is a group difference and the null model assuming that gene expression is explained by technical and within-patient variability.

Full model: gene count ~ offset(natural logarithm of effective library size) + group + (1|patient)

Null model: gene count ~ offset(natural logarithm of effective library size) + (1|patient)

We summarized our simulation setup in **Table 1**. For the final simulation run, the numbers of repeats/random data sets per condition was increased to 1,000.

Software

All analyses were conducted in R 4.1.0 [49], Bioconductor version 3.14. Research scripts are enclosed with this article (**Data S2**). Spatial transcriptomics data was handled with Seurat v4.1.1 [50]. Figures were made with ggplot2 [51] and viridis [52].

Data availability

Bulk RNA-sequencing GSE193066 [23] and spatial transcriptomics [34-36] were publicly available.

Results

Study endpoints

Sample size calculation requires definition of the study endpoint. We used a combination of historic bulk RNA-sequencing (GSE193066 [23]) and spatial transcriptomics data [34-36] to identify spatial gene expression pattens in baseline biopsies that correlated with fibrosis outcome in NAFLD.

At baseline, 108 genes had higher expression in fibrosis regressors and lower expression in fibrosis progressors compared to stable fibrosis individuals. Among these genes, 44 (40.7%) were associated with different aspects of lipid metabolism. By contrast, 187 genes had higher baseline expression in fibrosis progressors and lower baseline expression in fibrosis regressors. Among these genes, 89 (47.6%) were associated with actin cytoskeleton, cell projections such as cell leading edge and RHO GTPase activity. We used umbrella terms "lipid metabolism" and "cytoskeleton" to denote these 44 and 89 genes (**Data S3**).

9

The identified gene sets were aggregated into summary scores. Analysis of the summary scores confirmed that fibrosis progressors had lower baseline expression of lipid metabolism markers (**Figure 1A**) and higher expression of cytoskeleton markers (**Figure 1B**) irrespective of their disease severity at baseline.

Lipid metabolism score correlated with hepatocyte score (**Figure 1C**) while cytoskeleton score correlated with extracellular matrix (ECM) (**Figure 1D**) and mesenchyme scores (**Figure 1E**) in bulk NAFLD livers. Lipid metabolism score co-localized with the hepatocyte fraction in spatial transcriptomics on both normal and diseased livers. By contrast, cytoskeleton score co-localized with fibrosis areas marked by high expression of ECM and mesenchyme scores in diseased livers (**Figure 1F**). We could not further deconvolute which cell type of the mesenchymal lineage (stellate cells, activated stellate cells, vascular smooth muscle cells or fibroblasts) could lead to co-localization of this expression signal.

Based on this initial analysis, we formulated two study hypotheses:

1. Fibrosis regressors have higher baseline expression of the lipid metabolism score in hepatocytes than stable-fibrosis individuals.
2. Fibrosis progressors have higher baseline expression of the cytoskeleton score in the fibrotic niche than the fibrosis-stable patients

To facilitate reading, the study hypotheses were formulated relative to the patient group with higher expression. However, the differences in lipid metabolism and cytoskeleton scores were observed in both directions (i.e., two sided test, 1. – higher in fibrosis regressors/lower in progressors, 2. – higher in progressors/lower in regressors) and had approximately same magnitude compared to the fibrosis-stable group. The corresponding null hypotheses were that there were no group differences.

Hypothesis 1 was prioritized as the primary endpoint because we considered isolation of fibrotic niche more technically challenging and more likely to fail. However, both hypotheses were biologically plausible [53-55], and sample sizes were calculated to address both hypotheses.

Next, we needed to turn these hypotheses into quantifiable study endpoints. Hepatocytes are the most abundant fraction of liver tissue even in individuals with advanced fibrosis [38-40], so, effect size within the hepatocyte fraction would be expected to be only slightly different from the effect size observed in bulk liver. By contrast, fibrosis areas constitute 2.8% (2.1– 3.6%) in F1, 4.3% (3.3 - 5.4%) in F2, 4.8% (3.7 -7.4%) in F3 and 12.3% (8.4 -18.5%) in F4 NAFLD fibrosis stages [37]. Thus, expression signal originating within the fibrotic niche may be "diluted" in bulk liver, and we needed to translate the effect sizes between bulk liver and fibrotic niche and hepatocyte fraction. Individual genes that contributed to the summary scores did not always match the spatial expression pattern for the whole score. We found two representative markers PON1 and FLNA that had the most similar spatial expression pattern to the lipid metabolism score and cytoskeleton score, respectively (**Figure 2**) and based the subsequent calculations on these two genes. Based on the calibration curve analysis (**Figure S2**), we formulated the primary and secondary endpoints as:

1. Fibrosis regressors have 0.42±0.02 log2 fold higher baseline expression of PON1 in hepatocytes than stable fibrosis individuals.
2. Fibrosis progressors have 1.5±0.11 log2 fold higher baseline expression of FLNA in the fibrotic niche than stable fibrosis individuals.

Initial simulation study

Using the NanoString GeoMx® Human Whole Transcriptome Atlas Digital Spatial Profiler, gene expression data is typically collected from 6-12 regions of illumination (ROIs) per tissue sample (MAN-10108-01_GeoMx_DSP_Experimental_Design_Guideline.pdf from www.nanostring.com/GeoMxDSP). Historic data indicated that tissue samples may have uneven thickness and cell density may be variable within tissue sample. So, the expression measurements were expected to have large technical variability in addition to biological variability of expression for PON1 in hepatocytes and FLNA in the fibrotic niche.

Given that the outcome of the experiment may be affected by the sampling variability within the tissue sample, we conducted a simulation study to accurately mirror this sampling variability and estimate the required number of patients, ROI diameter and number of ROIs per patient.

11

Interestingly, the three methods applied head-to-head on the same synthetic data resulted in distinct power surfaces (**Figure 3** and **4**). However, in all cases, power to detect expression differences between the groups increased with increasing number of patients per group but not necessarily with increasing number of ROIs per patient (visible in Figure 3 and 4 in each column corresponding to a fixed N patients).

Final simulation study

The initial simulation study helped us to narrow down the range of experimental conditions to the largest ROI size 165μm, two ROIs per patient and at least 4 patients per group. We repeated the simulation study but with 1,000 synthetic data sets per condition (**Table 2**). All three methods were run on the same synthetic data, i.e., PON1 and FLNA read counts used to evaluate GeoDiff and glmer.nb and GLMMadaptive LRT methods were exactly the same. GeoDiff overestimated log2 fold change for the primary endpoint and underestimated log2 fold change for the secondary endpoint relative to the actual spike-in. Glmer.nb and GLMMadaptive LRT resulted in smaller log2 fold change estimates in the synthetic NanoString data sets than the actual spike-in but were close for the primary endpoint (PON1 in hepatocyte fraction median log2 FC 0.33 in simulation vs 0.42 in calibration curve analysis). Glmer.nb and GLMMadaptive LRT were faster than GeoDiff (e.g., final simulation for the secondary endpoint GeoDff 16.4 h, glmer.nb 6.6h and GLMMadaptive LRT 2.4h). However, we did not optimize the code, so runtimes could be reduced further. Convergence rate was 100% for glmer.nb and GeoDiff. GLMMadaptive LRT converged in 732 to 979 out of 1,000 iterations with higher failure rate on fewer patients.

Taking all these factors into consideration, the final sample size was based on the glmer.nb method. We calculated that we would need (10 patients x 2 hepatocyte ROIs + 5 patients (subset of the ten) x 2 fibrotic niche ROIs) x 3 experimental groups (stable fibrosis, fibrosis regressors, fibrosis progressors) = 90 samples subjected to sequencing at baseline. If paired follow-up biopsies are profiled for exploratory analysis from the same patients and for the same tissue areas, then the total sample size for baseline and equal number of follow-up samples is 180 samples.

12

Sensitivity analysis

The summary scores calculated from normalized data could also be approximated as sum of raw gene counts for all genes constituting the score. For example, Spearman correlation between lipid metabolism score calculated on normalized counts and sum of raw counts for lipid metabolism genes was median 0.79 (IQR 0.72 - 0.82) in the spatial transcriptomics samples. Sensitivity analysis using sum of gene counts resulted in similar fold change estimates in calibration curve analysis (**Figure S3**) and similar sample size estimates (**Table S1**). Fourteen and four patients per group were sufficient to reach 80% power for testing group differences in expression of the lipid metabolism and cytoskeleton genes in the hepatocyte fraction and fibrotic niche, respectively.

Discussion

Sample size calculation is a crucial step in planning studies that are appropriately powered to answer the research question. Limited methodological research is available for sample size calculation for spatial transcriptomics in clinical research. In this study, we outline how to gather required inputs (e.g. estimate effect size based on historic data) and perform sample size estimation for comparison between patient groups. Our biological question of interest was the difference in hepatic gene expression levels between patient groups (fibrosis progressors, fibrosis regressors and stable fibrosis patients) assayed with NanoString GeoMx spatial transcriptomics. The sampling variability within the tissue had to be accounted for but was not of primary interest. The three mixed effect model methods produced distinct power surfaces in the initial simulation run despite having similar conceptual basis. We attribute these differences to nuances of the software implementation pertaining to calculation of model likelihood and p-values. Still, with all three methods, power depended on the number of patients per group and not ROIs per patient, which is consistent with theoretical considerations [56]. The power of the study depends on the hypothesis/model parameter being tested (in this case, fixed effect regression coefficient for patient group). The number of ROIs per patient affects the ability to accurately estimate the random effects (in this case, patient-specific intercept terms) and not the fixed effect regression coefficients [56]. This represents sharp contrast to the current literature that focuses primarily on spatial variability of expression within the tissue and, accordingly, increasing the number

13

of spots or ROIs per tissue sample [5-7]. Whilst group differences between patients could be addressed with conventional experimental methods such as RNA-sequencing or even qPCR on FACS sorted or laser micro dissected tissue, this presents an impractical barrier in human research when spatial transcriptomics with the NanoString platform enables reliable quantification of gene expression levels from small areas of archived formalin fixed and paraffin embedded tissue.

Due to low amount of input material, there is a risk that some ROIs will fail technical quality control after acquisition of sequencing data. Therefore, we started the simulation with 2 ROIs per patient and endpoint. If one ROI fails technical quality control in a subset of individuals, the remaining replicates should still be sufficient to address the study endpoint. For example, if FLNA expression in 165μm fibrotic niche ROIs was compared between 5 fibrosis progressors and 5 fibrosis-stable individuals (5x2x2 ROI = 20 ROIs) and a third (6) of the ROIs were removed ("failed quality control") but in a way that did not result in complete loss of the affected patients, then power to detect the group difference with glmer.nb still remained 78.9% based on evaluation of 1,000 synthetic data sets.

glmer.nb yielded similar log2 fold change estimates to the actual spike-in, had good convergence rate and short runtime, so we decided to test primary and secondary endpoints with this method. We set the alpha level to 0.025 (0.05/2 endpoints). We decided that the primary endpoint will be tested first, the secondary endpoint will be tested next and any other analyses such as transcriptome-wide differential expression will be unpowered exploratory. Standard deviations of random intercepts were generally low: median 0.3 (IQR 0.1-0.5) for FLNA in fibrotic niche and median 0.1 (IQR 0.05-0.2) for PON1 in hepatocytes for all N patients and ROI size 165μm in the final simulation study. Type I error in this scenario should be close to the theoretical level [47] and we did not implement any additional steps to control it. However, sensitivity analysis could be conducted via bootstrap method [47]. In addition, expression of summary scores for the lipid metabolism and cytoskeleton markers could be compared between fibrosis progressors, fibrosis regressors and stable fibrosis individuals in sensitivity analysis. We considered this an acceptable approach. If transcriptome-wide significance level was desired, alpha could be set to 1.25e-05, which corresponds to (0.05/2 endpoints)/ca. 2,000 reliably detected genes with 165μm ROI in our reference data. At alpha 1.25e-05, 27 patients per

group would be required to achieve 80% power for the primary endpoint (PON1 in hepatocytes) and 15 patients per group would be required for the secondary endpoint (FLNA in fibrotic niche).

Our sample size calculation is subject to several other considerations:

First, the 58 patients with repeated biopsies in Fujiwara *et al*. cohort were at high risk to develop hepatocellular carcinoma and already had detectable liver fibrosis at baseline [23]. Therefore, the hypotheses formulated in this cohort might not generalize to all-comer NAFLD population.

Second, the initial study hypotheses (scores) were based on non-overlapping gene sets. We assumed that the two endpoints pertained to different aspects of biology and different tissue fractions and could therefore be considered independent, which may not necessarily be the case.

Third, NAFLD Brunt fibrosis score reflects the relative amount of fibrotic tissue area within the liver. We assumed that if the amount of fibrotic tissue subjected to sequencing (i.e., the ROI size) is fixed, then we do not need to adjust for fibrosis stage any longer. However, this assumption may not hold true if microscopic structure (e.g., cell composition or cell density) differs within fibrotic niche in patients with different fibrosis severity.

Finally, we focused on the sample size calculation and left other experimental considerations outside of the scope of the current study. For example, identification of ROIs (antibody selection, labelling with fluorescent dyes, reagent concentration etc) and sample randomization (the instrument has throughput of 4-8 slides a day, "GeoMx® Digital Spatial Profiler (DSP) Project Design Guide" document from www.nanostring.com/GeoMxDSP) are separate topics.

In conclusion, NanoString GeoMx is an emerging spatial transcriptomics technology. While some considerations such as definition of regions of interest and their size are technology-specific, we believe that the sample size calculations for spatial transcriptomics, as for any other technology, are predominantly shaped by the biological or clinical study hypothesis. The study hypothesis helps to define the analysis method (statistical model), the associated assumptions and the expected effect size or minimal clinically meaningful difference. In the absence of knowledge-based hypothesis or a dedicated pilot spatial transcriptomics study, the study endpoints can be inferred from prior bulk

RNA-sequencing data and re-scaled to tissue fractions corresponding to ROIs in spatial transcriptomics as illustrated in our case study.

Author contribution

V.A. defined the biological question and guided the study design. M.R. designed and conducted computational analysis and wrote the initial manuscript draft. Both authors revised and approved the final manuscript version.

Acknowledgments

We would like to thank Magnus Söderberg for valuable discussions on liver histology and Stephanie Ling and Scott Hoffmann for discussions around imaging aspects of data generation.

Funding

None.

About the authors

Maria Ryaboshapkina, M.Sc., is a senior scientist at AstraZeneca focusing on applied bioinformatics research in cardiovascular, renal and metabolic diseases.

Vian Azzu, MD PhD, is a principal scientist at AstraZeneca and practicing hepatology clinician in the UK National Health Service. Her research is focused on liver/metabolic diseases and improving patient health outcomes.

Conflict of interest

The authors are employees of AstraZeneca.

Tables

Table 1. Setup of the initial simulation.

Parameter	Parameter values
N patients per group	2-20
N ROIs per patient	2-10
PON1 spike-in 55µm hepatocyte Visium spots prior to aggregation	r0.43
FLNA spike-in 55µm fibrotic niche Visium spots prior to aggregation	r2.9
ROI diameters	80, 110, 165µm
N Visium spots aggregated to achieve the ROI diameters	2, 4, 9
Alpha	0.025
Power for primary endpoint (PON1 in hepatocyte fraction)	80%
Power for secondary endpoint (FLNA in fibrotic niche)	80%
N repeats per condition (endpoint x N patients x N ROIs x ROI diameter)	100
Analysis methods	GeoDiff, glmer.nb, GLMMadaptive LRT

Table 2. Results of the final simulation. (A) Primary endpoint PON1 in 165μm hepatocyte ROIs and 2 ROIs per patient. (B) Secondary endpoint FLNA in 165μm in fibrotic niche and 2 ROIs per patient. Number of patients is indicated per group. Log2 fold change is indicated between spiked-in and unmodified simulated data across 1,000 iterations per condition.

A.	GeoDiff		glmer.nb		GLMMadaptive	LRT
N patients	% Power	median (IQR) log2FC	% Power	median (IQR) log2FC	% Power	median (IQR) log2FC
4	27.3	1.08 (0.82-1.37)	51	0.33 (0.22-0.44)	38.4	0.32 (0.21-0.44)
5	32.8	1.04 (0.76-1.29)	56.2	0.33 (0.23-0.44)	47.8	0.33 (0.23-0.43)
6	35.7	0.97 (0.68-1.24)	61.9	0.33 (0.23-0.42)	54.0	0.33 (0.23-0.42)
7	36.2	0.88 (0.6-1.18)	72.3	0.33 (0.25-0.42)	63.8	0.34 (0.25-0.42)
8	37.8	0.85 (0.59-1.13)	74.5	0.33 (0.26-0.41)	68.7	0.33 (0.25-0.41)
9	42.9	0.82 (0.55-1.11)	79.4	0.33 (0.26-0.4)	73.9	0.33 (0.26-0.4)
10	44.3	0.78 (0.53-1.06)	82.5	0.33 (0.26-0.4)	78.2	0.33 (0.26-0.4)
11	47.3	0.78 (0.52-1.06)	86.4	0.33 (0.25-0.4)	82.8	0.33 (0.26-0.4)
12	47.2	0.74 (0.5-0.98)	89	0.33 (0.27-0.39)	86.3	0.33 (0.27-0.39)
13	49.3	0.74 (0.49-0.99)	90.7	0.33 (0.27-0.39)	88.4	0.33 (0.27-0.39)
14	50	0.73 (0.48-0.94)	93.1	0.33 (0.27-0.39)	91.5	0.33 (0.27-0.39)
15	49.3	0.68 (0.45-0.91)	95.1	0.33 (0.27-0.39)	93.6	0.33 (0.27-0.39)
16	49.4	0.64 (0.42-0.87)	96.6	0.33 (0.28-0.39)	95.2	0.33 (0.28-0.39)
17	56.6	0.68 (0.47-0.91)	97.1	0.33 (0.28-0.39)	96.2	0.33 (0.28-0.39)
18	49.5	0.6 (0.42-0.84)	97.4	0.33 (0.28-0.39)	96.6	0.33 (0.28-0.39)
19	56.8	0.64 (0.43-0.87)	98.9	0.34 (0.29-0.38)	98.4	0.34 (0.29-0.39)
20	55.5	0.61 (0.41-0.84)	99	0.33 (0.28-0.38)	98.8	0.33 (0.28-0.38)
B.	**GeoDiff**		**glmer.nb**		**GLMMadaptive**	**LRT**
4	33.4	0.83 (0.56-1.13)	72.7	1.04 (0.8-1.22)	60.4	1.02 (0.77-1.21)
5	36.9	0.82 (0.58-1.07)	80.8	1.04 (0.85-1.22)	70,9	1.03 (0.83-1.2)
6	39.9	0.82 (0.57-1.09)	84	1.04 (0.84-1.22)	76.2	1.04 (0.83-1.21)
7	43.4	0.81 (0.57-1.08)	87.7	1.05 (0.87-1.2)	82.0	1.05 (0.86-1.19)
8	46.2	0.81 (0.59-1.03)	91	1.03 (0.85-1.19)	86.6	1.02 (0.84-1.17)
9	49.7	0.81 (0.61-1.07)	93.7	1.04 (0.88-1.18)	90.7	1.03 (0.87-1.17)
10	54.9	0.81 (0.61-1.07)	93.5	1.03 (0.88-1.16)	91.4	1.02 (0.88-1.16)
11	56.7	0.84 (0.63-1.09)	97.2	1.04 (0.89-1.16)	95.8	1.04 (0.89-1.16)
12	55.9	0.8 (0.61-1.02)	97.5	1.04 (0.89-1.16)	96.3	1.03 (0.89-1.15)
13	59.7	0.82 (0.63-1.06)	98.1	1.02 (0.89-1.15)	97.3	1.02 (0.88-1.14)
14	58.6	0.8 (0.62-1.03)	98.5	1.03 (0.9-1.15)	98.2	1.03 (0.9-1.15)
15	64.3	0.81 (0.63-1.03)	99.5	1.02 (0.9-1.14)	99.3	1.02 (0.9-1.14)
16	66.3	0.82 (0.63-1.05)	99.3	1.04 (0.92-1.15)	98.9	1.04 (0.92-1.15)
17	64	0.82 (0.64-1.04)	99.5	1.03 (0.91-1.14)	99.3	1.03 (0.9-1.14)
18	67.8	0.8 (0.62-0.99)	99.8	1.02 (0.91-1.12)	99.8	1.02 (0.91-1.13)
19	68.3	0.81 (0.64-1.03)	100	1.03 (0.92-1.13)	100.0	1.04 (0.91-1.13)
20	69.3	0.8 (0.64-1.02)	99.8	1.03 (0.92-1.12)	99.7	1.03 (0.92-1.12)

Figure legends

Figure 1. Initial hypothesis generation. (A) Relative expression of lipid metabolism score in baseline biopsies stratified by 2-year outcome and baseline fibrosis stage in bulk NAFLD liver GSE193066. N stable-fibrosis individuals = 28, fibrosis progressors = 15 and fibrosis regressors = 15. Table on the side shows coefficients from the multivariate linear regression model, in which relative expression score is a function of 2-year outcome, fibrosis stage and NAS score at baseline. (B) Relative expression of cytoskeleton score in baseline biopsies stratified by 2-year outcome and baseline fibrosis stage in bulk NAFLD liver GSE193066. Notations are as in panel (A). Possible source of expression signal (C-E) in NAFLD bulk liver GSE193066 and (F) in spatial transcriptomics experiments. Correlations in spatial transcriptomics were computed across the total tissue area within each sample separately.

Figure 2. PON1 and FLNA as representative markers. Correlation between (A) PON1 and hepatocyte score and (B) FLNA and cytoskeleton score in bulk NAFLD liver GSE193066. Co-localization of (C) PON1 with lipid metabolism score in hepatocyte fraction and (D) FLNA with cytoskeleton score in fibrotic niche in representative non-tumour liver region adjacent to hepatocellular carcinoma. Colors indicate relative expression of summary marker scores. These scores are in arbitrary log2 units: blue – low, green – moderate, red – high.

Figure 3. Power surface obtained for the primary endpoint (PON1 in hepatocytes) in the initial simulation. This analysis was used to narrow down the range of experimental conditions (N patients x N ROIs per patient x ROI size).

Figure 4. Power surface obtained for the secondary endpoint (FLNA in fibrotic niche) in the initial simulation. This analysis was used to narrow down the range of experimental conditions (N patients x N ROIs per patient x ROI size).

References

1.   Villacampa EG, Larsson L, Mirzazadeh R, et al. Genome-wide spatial expression profiling in formalin-fixed tissues. *Cell Genomics* 2021;1:100065. doi: 10.1016/j.xgen.2021.100065.

2.   Merritt CR, Ong GT, Church SE, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* 2020;38:586-599. doi: 10.1038/s41587-020-0472-9.

3.   Roberts K, Aivazidis A, Kleshchevnikov V, et al. Transcriptome-wide spatial RNA profiling maps the cellular architecture of the developing human neocortex. *bioRxiv* 2021. doi: 10.1101/2021.03.20.436265.

4.   Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* 2022;19:534-546. doi: 10.1038/s41592-022-01409-2.

5.   Vickovic S, Schapiro D, Carlberg K, et al. Three-dimensional spatial transcriptomics uncovers cell type localizations in the human rheumatoid arthritis synovium. *Commun Biol* 2022;5:129. doi: 10.1038/s42003-022-03050-3.

6.   Baker EAG, Schapiro D, Dumitrascu B, et al. Power analysis for spatial omics. *bioRxiv* 2022. doi: 10.1101/2022.01.26.477748.

7.   Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods* 2020;17:193-200. doi: 10.1038/s41592-019-0701-7.

8.   Noll JM, Augello CJ, Kurum E, et al. Spatial Analysis of Neural Cell Proteomic Profiles Following Ischemic Stroke in Mice Using High-Plex Digital Spatial Profiling. *Mol Neurobiol* 2022;59:7236-7252. doi: 10.1007/s12035-022-03031-x.

9.   Keam SP, Halse H, Nguyen T, et al. High dose-rate brachytherapy of localized prostate cancer converts tumors from cold to hot. *J Immunother Cancer* 2020;8. doi: 10.1136/jitc-2020-000792.

10.   Younossi ZM, Koenig AB, Abdelatif D, et al. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73-84. doi: 10.1002/hep.28431.

11.   Ekstedt M, Nasr P, Kechagias S. Natural History of NAFLD/NASH. *Curr Hepatol Rep* 2017;16:391-397. doi: 10.1007/s11901-017-0378-2.

12.   Yki-Jarvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol* 2014;2:901-910. doi: 10.1016/S2213-8587(14)70032-4.

13.   Pais R, Charlotte F, Fedchuk L, et al. A systematic review of follow-up biopsies reveals disease progression in patients with non-alcoholic fatty liver. *J Hepatol* 2013;59:550-556. doi: 10.1016/j.jhep.2013.04.027.

14.   McPherson S, Hardy T, Henderson E, et al. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *J Hepatol* 2015;62:1148-1155. doi: 10.1016/j.jhep.2014.11.034.

15.   Ekstedt M, Hagstrom H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015;61:1547-1554. doi: 10.1002/hep.27368.

16.   Hagstrom H, Nasr P, Ekstedt M, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. *J Hepatol* 2017;67:1265-1273. doi: 10.1016/j.jhep.2017.07.027.

17.   Adams LA, Sanderson S, Lindor KD, et al. The histological course of nonalcoholic fatty liver disease: a longitudinal study of 103 patients with sequential liver biopsies. *J Hepatol* 2005;42:132-138. doi: 10.1016/j.jhep.2004.09.012.

18.   Lin TY, Yeh ML, Huang CF, et al. Disease progression of nonalcoholic steatohepatitis in Taiwanese patients: a longitudinal study of paired liver biopsies. *Eur J Gastroenterol Hepatol* 2019;31:224-229. doi: 10.1097/MEG.0000000000001285.

19.   Pelusi S, Petta S, Rosso C, et al. Renin-Angiotensin System Inhibitors, Type 2 Diabetes and Fibrosis Progression: An Observational Study in Patients with Nonalcoholic Fatty Liver Disease. *PLoS One* 2016;11:e0163069. doi: 10.1371/journal.pone.0163069.

20.   Liu YL, Reeves HL, Burt AD, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat Commun* 2014;5:4309. doi: 10.1038/ncomms5309.

21.   Mancina RM, Dongiovanni P, Petta S, et al. The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. *Gastroenterology* 2016;150:1219-1230 e1216. doi: 10.1053/j.gastro.2016.01.032.

22.   Qian T, Fujiwara N, Koneru B, et al. Molecular Signature Predictive of Long-Term Liver Fibrosis Progression to Inform Antifibrotic Drug Development. *Gastroenterology* 2022;162:1210-1225. doi: 10.1053/j.gastro.2021.12.250.

23.   Fujiwara N, Kubota N, Crouchet E, et al. Molecular signatures of long-term hepatocellular carcinoma risk in nonalcoholic fatty liver disease. *Sci Transl Med* 2022;14:eabo4474. doi: 10.1126/scitranslmed.abo4474.

24.   Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. doi: 10.1186/s13059-014-0550-8.

25.   Govaere O, Cockell S, Tiniakos D, et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med* 2020;12. doi: 10.1126/scitranslmed.aba4448.

26.   Carbon S, Ireland A, Mungall CJ, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009;25:288-289. doi: 10.1093/bioinformatics/btn615.

27.   Thomas PD, Ebert D, Muruganujan A, et al. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* 2022;31:8-22. doi: 10.1002/pro.4218.

28.   Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25-29. doi: 10.1038/75556.

29.   Gene Ontology C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* 2021;49:D325-D334. doi: 10.1093/nar/gkaa1113.

30.   Gillespie M, Jassal B, Stephan R, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res* 2022;50:D687-D692. doi: 10.1093/nar/gkab1028.

31.   Kuppe C, Ibrahim MM, Kranz J, et al. Decoding myofibroblast origins in human kidney fibrosis. *Nature* 2021;589:281-286. doi: 10.1038/s41586-020-2941-1.

32.   Ramachandran P, Dobie R, Wilson-Kanamori JR, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 2019;575:512-518. doi: 10.1038/s41586-019-1631-3.

33.   Brosch M, Kattler K, Herrmann A, et al. Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. *Nat Commun* 2018;9:4150. doi: 10.1038/s41467-018-06611-5.

34.   Wu R, Guo W, Qiu X, et al. Comprehensive analysis of spatial architecture in primary liver cancer. *Sci Adv* 2021;7:eabg3750. doi: 10.1126/sciadv.abg3750.

35.   Andrews TS, Atif J, Liu JC, et al. Single-Cell, Single-Nucleus, and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal Heterogeneity. *Hepatol Commun* 2022;6:821-840. doi: 10.1002/hep4.1854.

36.   Guilliams M, Bonnardel J, Haest B, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* 2022;185:379-396 e338. doi: 10.1016/j.cell.2021.12.018.

37.   Masugi Y, Abe T, Tsujikawa H, et al. Quantitative assessment of liver fibrosis reveals a nonlinear association with fibrosis stage in nonalcoholic fatty liver disease. *Hepatol Commun* 2018;2:58-68. doi: 10.1002/hep4.1121.

38.   Racanelli V, Rehermann B. The liver as an immunological organ. *Hepatology* 2006;43:S54-62. doi: 10.1002/hep.21060.

39.   Bogdanos DP, Gao B, Gershwin ME. Liver immunology. *Compr Physiol* 2013;3:567-598. doi: 10.1002/cphy.c120011.

40.   Vertemati M, Minola E, Goffredi M, et al. Computerized morphometry of the cirrhotic liver: comparative analysis in primary biliary cirrhosis, alcoholic cirrhosis, and posthepatitic cirrhosis. *Microsc Res Tech* 2004;65:113-121. doi: 10.1002/jemt.20110.

41.   Naoumov NV, Brees D, Loeffler J, et al. Digital pathology with artificial intelligence analyses provides greater insights into treatment-induced fibrosis regression in NASH. *J Hepatol* 2022;77:1399-1409. doi: 10.1016/j.jhep.2022.06.018.

42.   Pirhonen J, Arola J, Sadevirta S, et al. Continuous Grading of Early Fibrosis in NAFLD Using Label-Free Imaging: A Proof-of-Concept Study. *PLoS One* 2016;11:e0147804. doi: 10.1371/journal.pone.0147804.

43.   GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580-585. doi: 10.1038/ng.2653.

44.   Yang L, Yang Z, Danaher P, et al. Background modeling, Quality Control and Normalization for GeoMx RNA data with GeoDiff. *bioRxiv* 2022. doi: 10.1101/2022.05.26.493637.

45.   Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-140. doi: 10.1093/bioinformatics/btp616.

46.   Bates D, Mächler M, Bolker B, et al. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 2015;67:1 - 48. doi: 10.18637/jss.v067.i01.

47.   Tsonaka R, Spitali P. Negative Binomial mixed models estimated with the maximum likelihood method can be used for longitudinal RNAseq data. *Brief Bioinform* 2021;22. doi: 10.1093/bib/bbaa264.

48.   Rizopoulos D. GLMMadaptive: generalized linear mixed models using adaptive Gaussian quadrature. R package version 0.8-5. 2019.

49.   R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022.

50.   Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573-3587 e3529. doi: 10.1016/j.cell.2021.04.048.

51.   Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer, 2016.

52.   Garnier S, Ross N, Rudis B, et al. sjmgarnier/viridis: CRAN release v0.6.2. *Zenodo* 2021. doi: 10.5281/zenodo.5579397.

53.   Marsillach J, Camps J, Ferre N, et al. Paraoxonase-1 is related to inflammation, fibrosis and PPAR delta in experimental liver disease. *BMC Gastroenterol* 2009;9:3. doi: 10.1186/1471-230X-9-3.

54.   Cui X, Zhang X, Yin Q, et al. F□actin cytoskeleton reorganization is associated with hepatic stellate cell activation. *Mol Med Rep* 2014;9:1641-1647. doi: 10.3892/mmr.2014.2036.

55.   Hijazi N, Rockey DC, Shi Z. The cellular microenvironment and cytoskeletal actin dynamics in liver fibrogenesis. *Biocell* 2022;46:2003-2007. doi: 10.32604/biocell.2022.020171.

56.   Snijders T. Power and sample size in multilevel modeling. In: Everitt B. S., Howell D. C. (eds). Encyclopedia of Statistics in Behavioral Science. Chicester: Wiley, 2005, 1570–1573.
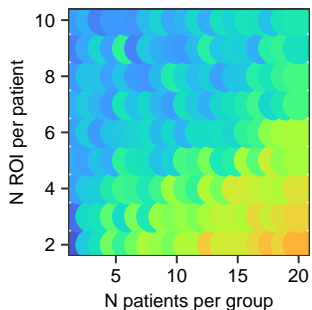
A

Lipid metabolism score, log2 units — Baseline fibrosis (F1, F2, F3)

fibrosis outcome: stable, regressor, progressor

	log2 FC	p value
regressor vs stable	0.33	1e−04
progressor vs stable	−0.35	2e−04
baseline fibrosis	−0.25	4e−05
baseline NAS	0	0.91

B

Cytoskeleton score, log2 units — Baseline fibrosis (F1, F2, F3)

	log2 FC	p value
regressor vs stable	−0.41	1e−05
progressor vs stable	0.39	1e−04
baseline fibrosis	0.36	2e−07
baseline NAS	0.01	0.6

C Correlation 0.53

Lipid metabolism score, log2 units vs Hepatocytes score, log2 units

D Correlation 0.68

Cytoskeleton score, log2 units vs ECM score, log2 units

E Correlation 0.67

Cytoskeleton score, log2 units vs Mesenchyme score, log2 units

F

lipid metabolism / cytoskeleton

Rows: Hepatocytes, ECM, Mesenchyme

Columns: Healthy, Healthy, Healthy, Steatotic, Steatotic, Steatotic, adj. to HCC, adj. to HCC, adj. to HCC, adj. to HCC, adj. to HCC

Correlation scale: 0.75, 0.50, 0.25, 0.00, −0.25

A Correlation 0.76

B Correlation 0.79

C

Hepatocytes

Lipid metabolism score

PON1

D

ECM

Cytoskeleton score

FLNA

GeoDiff · 80µm ROI — N patients per group vs N ROI per patient, Power %

glmer.nb · 80µm ROI — N patients per group vs N ROI per patient, Power %

GLMMadaptive LRT · 80µm ROI — N patients per group vs N ROI per patient, Power %

GeoDiff · 110µm ROI — N patients per group vs N ROI per patient, Power %

glmer.nb · 110µm ROI — N patients per group vs N ROI per patient, Power %

GLMMadaptive LRT · 110µm ROI — N patients per group vs N ROI per patient, Power %

GeoDiff · 165µm ROI — N patients per group vs N ROI per patient, Power %

glmer.nb · 165µm ROI — N patients per group vs N ROI per patient, Power %

GLMMadaptive LRT · 165µm ROI — N patients per group vs N ROI per patient, Power %