Bone marrow haematopoietic stem cells influence liver homeostatic networks and cancer development

Ghazal Alipour Talash[1#], Peter Langfelder[2#], Daniele Vitale[1#], Mahmoud Karimi Azardaryany[1], T. Grant Belgard[3], Jocelyn Choo[4], Geraint Rogers[4], Vikki Ho[1], Mehdi Ramezani-Moghadam[1], Suat Dervish[5], Joey Lai[5], Brian S. Gloss[5], Duncan McLeod[6], Mohammed Eslam[1], Chris Liddle[1], Liang Qiao[1], Jacob George[1*], Saeed Esmaili[1,7*]

[1]Storr Liver Centre, Westmead Institute for Medical Research, Westmead Hospital and University of Sydney, NSW, Australia
[2]Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA
[3]The Bioinformatics CRO, Orlando, Florida, USA.
[4]SAHMRI Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia.
[5]Westmead Research Hub, Westmead Institute for Medical Research, NSW, Australia
[6]Department of Anatomical Pathology, Westmead Hospital, Sydney, NSW, Australia
[7]Liver and Pancreatobiliary Diseases Research Center, Digestive Disease Research Institute, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
[#] Equal authors
***Correspondence:** Jacob.george@sydney.edu.au; saeed.esmaili@sydney.edu.au

Conflicts of interest: All authors disclose no conflicts.

The article contains recycled material.

ABSTRACT

In fatty liver disease, systemic homeostasis is perturbed. While pre-clinical models are used to understand its pathogenesis, translating this knowledge to patients is difficult. However, by focusing on the most preserved homeostasis systems between species and models, novel disease dimensions can be unearthed. We interrogated core liver gene co-expression networks in a mouse model of liver cancer following dietary challenge. Immunometabolic modules showed temporal changes under the influence of diet duration and aging. The behaviour of immune network in tumours mirrored their counterparts in non-tumour liver. A high immune response network was associated with a lower tumour burden in mice and humans. This module in mice was enriched for genes related to haematopoietic cell differentiation. Consistently, the bone marrow haematopoietic stem and progenitor cells response was reflective of the liver immune response. Linking haematopoiesis to hepatic homeostasis uncovers a hitherto unexplored dimension of tissue crosstalk that can inform pathogenesis.

KEYWORDS: HCC, fatty liver disease, core-liver homeostasis co-expression networks, tumour permissive microenvironment, immune response, liver-bone marrow cross talk, haematopoiesis, WGCNA

INTRODUCTION

The natural history of fatty liver disease and its extra-hepatic associations reflects multi-system pathology with many target organs. Hepatic complications include cirrhosis and liver cancer (hepatocellular cancer, [HCC]) though cancer can also arise in a fatty liver without cirrhosis (Mittal et al., 2016; Paradis et al., 2009). Hepatotropic viruses such as Hepatitis B and C are a major cause of liver cancer globally, but fatty liver disease is predicted to become the dominant aetiology in the next two decades (Younossi et al., 2018; Younossi et al., 2015). Considering that mutations occur in normal tissues including the liver (Yizhak et al., 2019), the mutational load *per se* is not the only requisite for cancer development (DeGregori, 2017) and alterations in the tissue microenvironment are equally pivotal in selecting for onco-primed cells. Less understood for liver cancer development is the complex interplay between hepatic and extra-hepatic homeostatic mechanisms. Considering these interactions at the systems level is pivotal to understanding the progression of fatty liver disease to cancer.

The approval of Atezolizumab with Bevacizumab as the standard of care for HCC (Finn et al., 2020) highlights the importance of targeting both the tumour and its microenvironment. Nevertheless, a recent analysis of clinical trials in HCC has shown that tumours of viral origin respond to immunotherapy better than to oral multi-kinase inhibitors, while those arising in the context of fatty liver do not (Pfister et al., 2021). This could perhaps be attributed to differences in the microenvironment between livers with viral infection and those with liver fat accumulation. Of interest, Spitzer et al., demonstrated that a systemic immune response to immunotherapy is pivotal for drug response (Spitzer et al., 2017). Teleologically, alterations in an organism's microenvironment with systemic homeostatic perturbations could also impact the liver. How this occurs in the context of fatty liver and liver cancer, including the interplay between the tumour ecosystem and local and systemic immune surveillance represents the best hope for understanding HCC biology for the development of new treatments.

Like any complex system, breaking the components to its constituent parts can provide insights on the function of the system *in toto*. These sub-components typically possess patterns that appear more often than by chance (Milo et al., 2002; Shen-Orr et al., 2002). Recently, using a systems approach we discovered preserved gene co-expression networks (modules) in pre-clinical dietary models of fatty liver disease, human fatty liver disease, and liver cancer. We called the preserved networks the core liver homeostatic co-expression networks (core-modules). These core modules demonstrated species- and disease-specific behaviours (Blencowe and Yang, 2021; Esmaili et al., 2021). In the current study we reasoned that scrutinising core network behaviours in relation to cancer burden after dietary challenge could provide novel insights. We performed systems analysis on liver and tumour tissues followed by preservation analysis with five available datasets: a mouse non-tumour dataset, human fatty liver, and the TCGA and ICGC (France, and Japan) liver cancer datasets. We replicated the previous findings but also observed temporal effects on network behaviour influenced by diet duration and ageing. Of particular interest, the hepatic immune response was associated with unique bone marrow haematopoietic stem cell responses. This crosstalk between liver and bone marrow uncovers a novel mechanism that influences liver cancer development.

RESULTS

Liver cancer development is impacted by diet composition

We used a mouse liver cancer model to evaluate hepatic phenotypic outcomes when challenged with diets of varying composition; we used this to define core liver homeostatic networks. To do this, we injected 14-day old male C57BL6 mice with the pro-carcinogen diethylnitrosamine (DEN). Starting at 10 weeks of age, mice were assigned to one of 4 diet groups for 26 weeks: A) normal chow (NC); B) high sucrose (HS); C) NC diet supplemented with cholic acid (CA); and D) HS diet with added cholesterol and cholic acid (HS_Chol2%_CA), hereafter referred to as a cholesterol rich diet (ChR) (Figure 1A). As we have reported, a ChR diet significantly alters liver homeostasis (Esmaili et al., 2021) and was therefore used in this tumour study. The ChR diet contains sucrose and cholic acid and hence the HS and CA diets were used as additional controls.

The body weights of mice fed the ChR and CA diets declined initially then remained steady through the experiment (Supplementary Figure 1A). Glucose tolerance was tested after 10 weeks on diet (week 20) and was found to be impaired in mice fed the HS diet (Figure 1B). At experiment end (week 36), mice fed the HS diet also had markedly greater body weights (Figure 1C). Both the ChR and CA groups had lower body weights than the control diet fed mice (Figure 1C), a higher liver/body weight ratio, and less epididymal and subcutaneous fat (Supplementary Figure 1B-E) as well as lower blood glucose levels (Supplementary Figure 1F). Higher blood triglyceride levels were observed in HS and CA groups while a higher blood cholesterol was noted in the HS and ChR groups (Supplementary Figure 1G-H).

All mice regardless of diet developed macroscopically visible tumours at 36 weeks. The CA diet group developed more and bigger tumours ($p < 0.05$) than NC and ChR groups, while we detected a higher trend in the HS group (Figure 1D-F). In contrast, mice on the ChR diet, despite being exposed to sucrose and cholic acid had a similar tumour burden as the NC group. We confirmed the findings by histology (Figure 1G). The HS group demonstrated steatosis ($P<0.01$), while the ChR group showed steatosis and inflammation (Figure 1G-H). We detected a trend to higher ALT with the CA and ChR diets and to lower ALT in the HS group (Figure 1I). In sum, the ChR diet reduced tumour burden despite liver injury and the presence of pro-tumorigenic factors such as sucrose and CA in the diet. Of note, the lower body weight and improved glucose metabolism cannot fully explain the reduced tumour burden in ChR diet fed mice as the CA group despite having similar weights, had an increase in liver cancer.

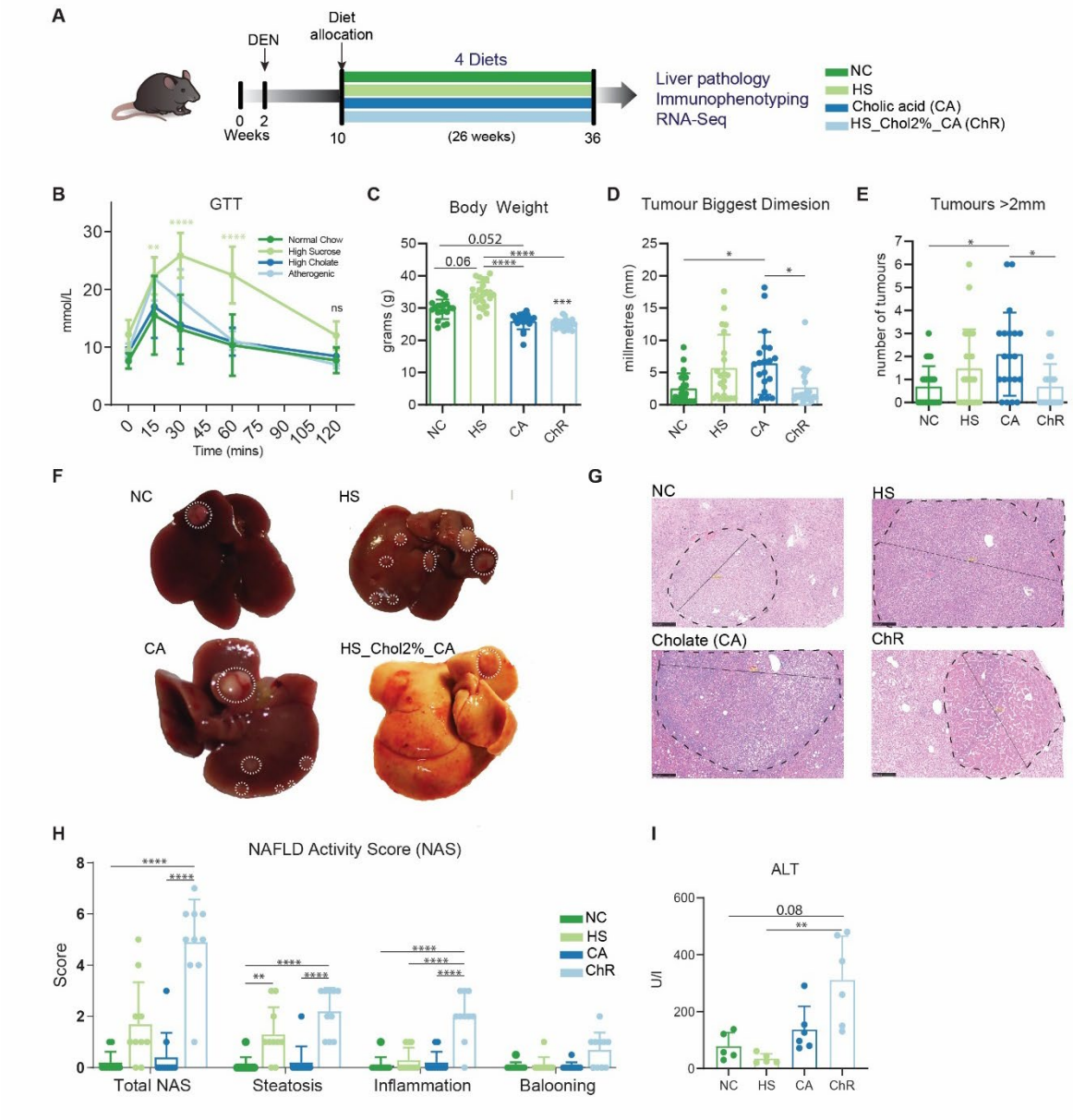

Figure 1. Experimental studies on liver tumourigenesis.
A) Diagram of mouse DEN tumour study. To induce liver tumours, DEN (25 mg/kg) was administered at the age of 2 weeks. At 10 weeks of age mice were allocated into four dietary groups: i) normal chow (NC); ii) high sucrose (HS), iii) a diet with added cholic acid (CA), or iv) a cholesterol rich diet (ChR) for 26 weeks. B) Glucose tolerance test (GTT) shows impairment in glucose homeostasis in the HS group. C) Body weight at experiment end (36 weeks). D-E) Tumour biggest dimension and multiplicity >2mm (number of tumours with maximum diameter above 2mm per mouse). F) Autopsy photos representing the tumour bearing livers of DEN-treated mice. G) Representative H&E stained liver sections of DEN-injected mice fed the diets as above. H) NAFLD activity score (NAS) of non-tumour area indicates higher hepatic steatosis in HS diet fed mice with steatosis, lobular inflammation and ballooning in the ChR diet. I) ALT levels are higher in the cholesterol rich diet group. Error bars represent mean ±1 standard deviation (SD). One-way ANOVA was used to test for significant differences in means with Dunn's (non-parametric) multiple comparison post-hoc test (Figs.1C,1D,1E, & 1I). Two-way ANOVA with Tukey's (parametric) multiple comparison post-hoc test was used in Figs.1B,and 1H . ****, ***, ** and * indicate P <0.0001, P<0.001, P<0.01 and P<0.05 respectively.

High concordance between differentially expressed genes in the DEN tumour study and a short-term diet study without DEN

To take an unbiased approach to liver cancer development, we performed RNA-Seq on tumour and non-tumour tissues of mice on the different diets. The number of significantly differentially expressed (DE) genes for all contrasts is shown in Figure 2A and supplementary Figure 1I-K. In in-between groups analysis, there were increasing numbers of DE genes at FDR<0.05 between non-tumour liver of NC vs. HS, vs CA and vs ChR (1342, 6383 and 8310 genes, respectively) (Figure 2A).

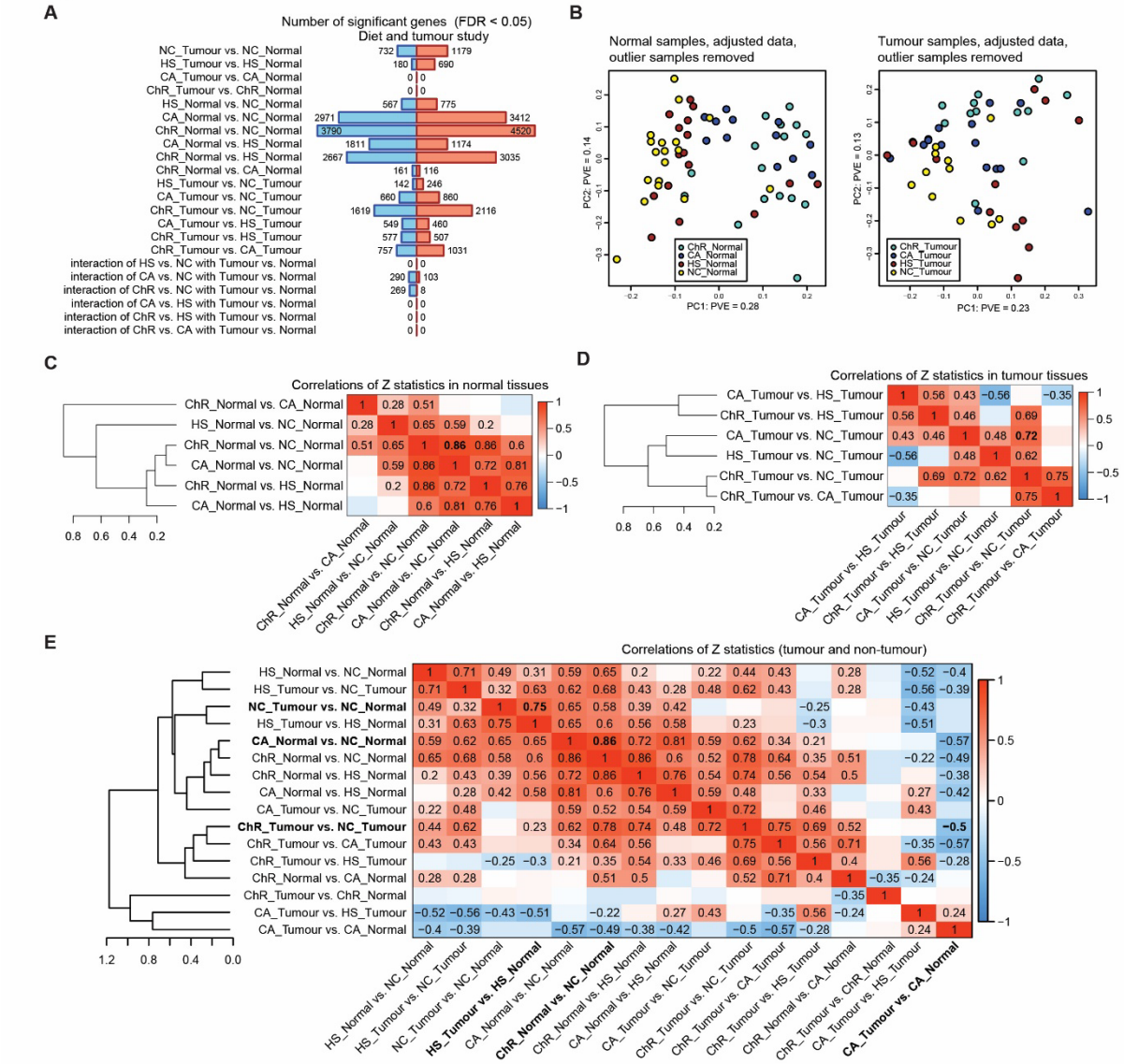

Figure 2. Comparisons of genome wide effects of diet and tumour formation on tumour and non-tumour liver transcriptome. A) Numbers of significantly (FDR < 0.05) differentially expressed genes for all tests. Blue (red) represent down- (up-)regulated genes. B) Principal component (PC) plot showing the 2 leading PCs after outlier removal. Each point represents one sample, colour indicates diet. C) Genome-wide correlations of Z statistics for all diet contrasts in normal tissues. D) Genome-wide correlations of Z statistics for all diet contrasts in tumour tissues. E-F) Genome-wide correlations of Z statistics for all diet and condition contrasts. Heatmap colours represent correlation; correlations whose absolute value is above 0.2 are shown explicitly.

For these same diet comparisons in the tumours, the number of DE genes were 388, 1520 and 3735 genes, respectively (Figure 2A). In within group analysis, the number of DE genes between tumour and non-tumour liver decreased in this order: NC (1911 genes), HS (870 genes), CA (0 genes), and ChR (0 genes) (Figure 2A). This data indicates that there is less of a distinction between tumour and non-tumour tissue in the CA and ChR diet fed mice.

Next, in principal component analysis (PCA) of non-tumour liver and tumours there was clustering of NC and HS samples, while ChR and CA diet groups mainly clustered together (Figure 2B). We then compared the genome-wide effects of diet composition and condition (non-tumour liver versus tumour) contrasts by correlating the vectors of Z statistics (Figure 2C-D). This indicated that the ChR vs. NC, and CA vs. NC in non-tumour liver and tumour samples are strongly correlated at 0.86 and 0.72, respectively. Thus, ChR and CA diets have a similar effect when compared to NC. In NC and HS diets, tumour vs. non-tumour liver DE is similar (Figure 2C-E, cor = 0.71). These correlations are very strong but potentially biased upward because they share a common subset of samples (e.g., Z for ChR vs. NC and Z for CA vs. NC have the level NC in common). Supplementary Figure 1L contains a heatmap in which, for clarity, the potentially biased correlations are set to zero. Overall, this data confirms that at the level of gene expression, NC and HS diets behave similarly as do the ChR and CA diets, with similar changes when comparing the tumour to the non-tumour liver in each diet comparison.

To understand the impact of mouse age, diet, tumour presence and their genome-wide effects on the transcriptome, we compared mice on diets 34 weeks after DEN injection to that of a previous study (8 weeks on diet and without DEN) (Esmaili et al., 2021). We correlated the vectors of DE Z statistics among these two studies (Figure 3A). Focusing on non-tumour livers, there were strong correlations between the effect of ChR vs. NC, CA vs. NC, and HS vs. NC in the current study and the ChR vs. NC in the previous study (cor = 0.78, cor= 0.69, core= 0.54, respectively, p<1e-200 for all) (Figure 3A, and supplementary Figure 2A). Unexpectedly, there was poor correlation (cor=0.22) between HS vs. NC in the current study and the same diet in the previous study without DEN which was of shorter (8 weeks) duration (Figure 3A). This indicates that long-term exposure to a HS diet and possibly the presence of tumour markedly impacts liver gene expression of HS diet fed mice.

Poor concordance between mouse, human fatty liver and human liver cancer at the level of gene expression

We next evaluated the concordance of DE genes in human liver (fatty liver and cancer) datasets with DE genes for non-tumour liver and tumour contrasts in mice on the 4 diets. We found weaker positive correlations (0.2 to 0.34) for the human tumours (Figure 3B-C and supplementary Figure 2B) as well as for human fatty liver (absolute values between 0.2 and 0.4) (Supplementary Figure 3A-B). These relatively weak correlations imply that transferring insights from mice tumour models to humans is challenging. Hence, we reasoned that a co-expression network analysis (WGCNA) (Zhang and Horvath, 2005a) approach would be more informative and allow us to define the most preserved co-expression modules (core modules) in relation to tumour burden *between* species.

We performed WGCNA analysis on variance stabilized data to reduce thousands of genes to a small number of transcriptionally coherent modules that represent distinct transcriptional responses, in this instance to diet and

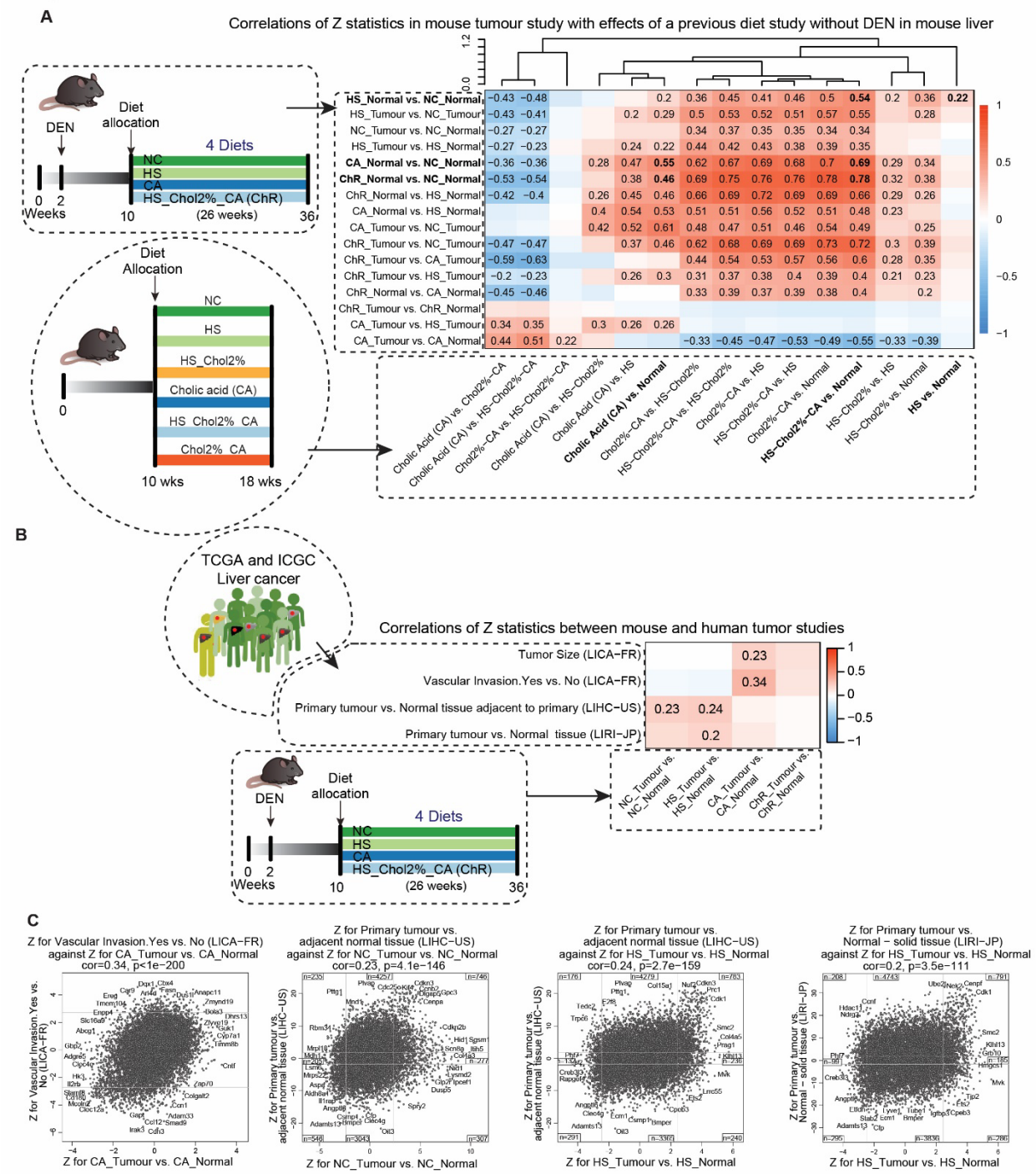

Figure 3. Comparisons of liver transcriptomes between mouse tumour study, a previous study of pre-clinical dietary models of fatty liver disease and human cancer datasets (TCGA, LICA-FR & LIRI-JP). A) Genome-wide correlations of Z statistics for all contrasts (diet and condition [tumour, non-tumour]) in this study (rows) vs. diet contrasts of pre-clinical dietary models of fatty liver disease (columns). The schematics depicts current tumour study and a previous diet study without DEN. B) Genome-wide correlations of Z statistics between the current study and human cancer TCGA and ICGC datasets (LICA-FR and LIRI-JP). Correlations were calculated for tumour size and vascular invasion in LICA-FR. C) Scatterplots of Z statistics for selected contrasts in the current study (x-axis) and human tumour datasets (y-axis). Each point represents a gene. Heatmap colors represent correlation; correlations whose absolute value is above 0.2 are shown explicitly. In scatter plots, grey lines represent approximate FDR=0.05 thresholds. Insets give the numbers of genes in each border region of the plot and hypergeometric p-values for the corner overlaps.

disease (tumour/ non-tumour liver) (Figure 4A, suppl Figure 4A-C, and suppl Figure 5).

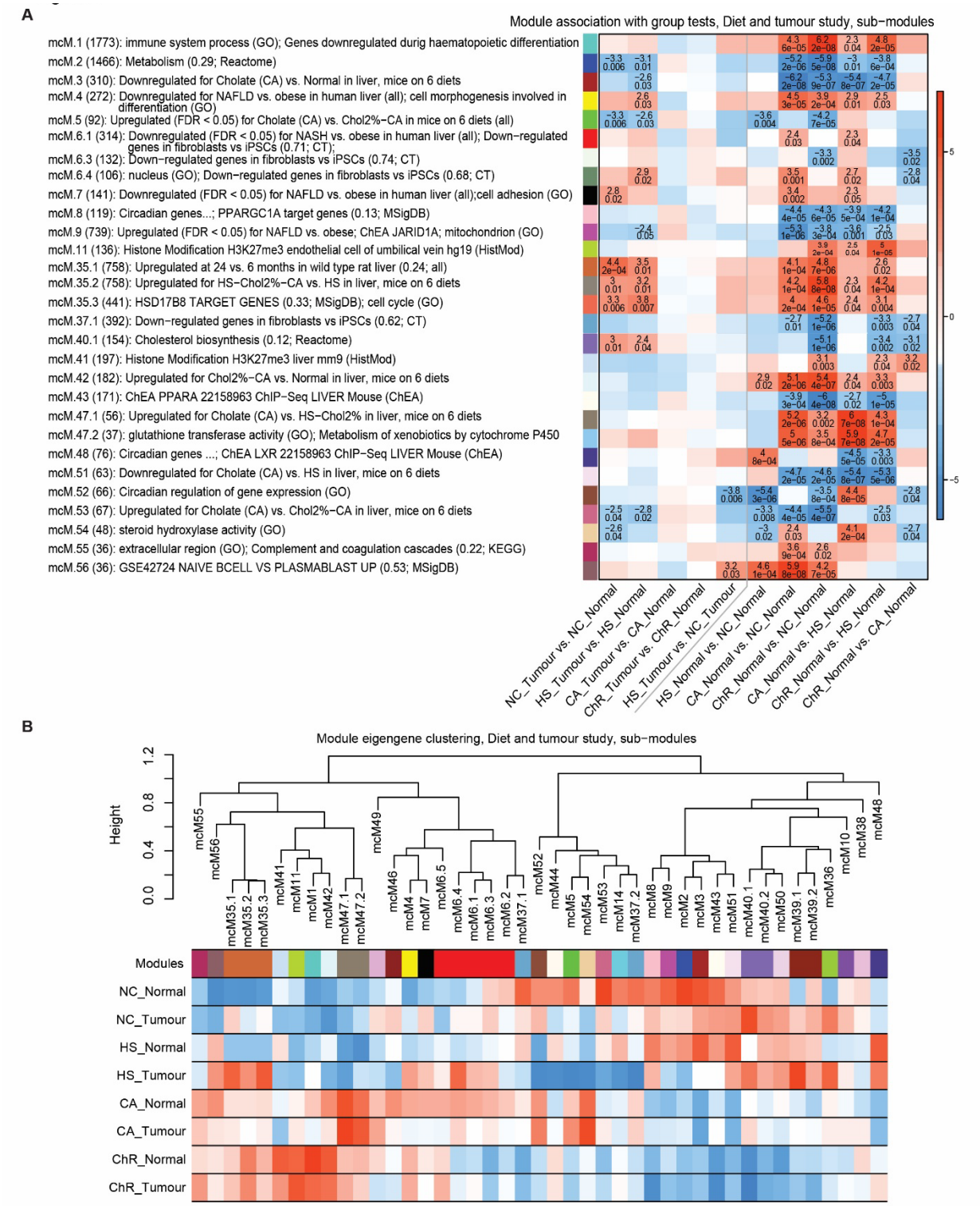

Figure 4. Co-expression network analysis of normal and tumour mouse liver samples. A) Association of selected module eigengenes of sub modules across all samples with diet and condition (tumour vs. normal) contrasts. Each row corresponds to one of the modules strongly associated with at least one contrast. Row labels indicate the numeric module label, module size and top enrichment terms. Colour rectangles correspond to the module colour in Supplementary Figure 4A. Each column corresponds to a diet or condition contrast. In the heatmap, numbers give the association significance Z statistic and the FDR estimate. For clarity, Z and FDR are only shown for those cells where FDR<0.05. Colour indicates strength and direction of the association Z statistic. B) Module clustering analysis of diet and tumour study.

To distinguish between modules in different datasets we used the following labelling strategy: modules in current mouse cancer (mc) study with DEN injection as mcM; modules in previous mouse 8 weeks diet study without DEN as (mM); modules in human fatty liver (hf) dataset as hfM; modules in human liver cancer (hc) datasets as hcM. Module labels were chosen so that modules that strongly overlap with modules from the previous 8 weeks study without DEN carry similar labels (e.g., mcM1 overlaps strongly with mM1 and mcM2 overlaps with mM2; both relate to immune response and metabolism and carry labels M1 and M2, respectively) (Supplementary Figure 6A-C). Since each co-expression module groups together correlated genes, each module can be represented by a single representative expression profile called a module eigengene (ME). The module eigengene is defined as the first singular vector of the standardized module expression matrix; this vector explains most of the variance of the module. Module eigengenes lead in this way to a natural measure of similarity (membership) of all individual genes to all modules. Genes with highest module membership are called hub genes.

We analysed the association of module eigengenes with diet and condition (tumour/ non-tumour liver) (Figure 4A). We find generally stronger associations between modules and diets in non-tumour samples (Figure 4A, columns 6 to 10). Relatively few modules are strongly associated with tumour vs. non-tumour liver status. The associations are strongest in NC and HS diets (Figure 4A, columns 1 and 2), while none of the modules are strongly associated with tumour vs. non-tumour liver in the CA and ChR diets (Figure 4A, columns 3 and 4). Figure 2B depicts module clustering of the current mouse diet and tumour study. This indicates that a combination of module responses possibly contributes to tumour outcome. Considering the poor concordance between the mouse tumour study and the human liver cancer data, it is possible that many of the modules in mice would not be preserved, and hence are of less translational value. Therefore, these findings suggest the need for a core-based strategy that focuses on the most preserved modules.

Temporal behaviour of immune response co-expression networks might influence tumour outcomes

Considering the weak concordance between DE genes in the mouse model of liver cancer and human liver tumours (Figure 3B) we reasoned that focusing on the most preserved modules (core modules) could have translational value. Consistent with our previous report (Esmaili et al., 2021), our reciprocal module preservation analysis demonstrates that the highest preserved modules between mice and human tumour datasets (US, FR, and JP) is the immune response (Figure 5A-D and Supplementary Figure 5A). In the LICA_FR dataset which includes phenotypic data such as tumour size and presence or absence of vascular invasion, the immune response module hcM1_LICA_FR is negatively correlated with tumour size (cor= -0.23, p=0.02; Figure 5E). Higher hcM1_LICA_FR is also associated with a lower rate of vascular invasion (Figure 5F). Thus, hcM1 (immune response) shows a pattern in which a higher M1 has protective effects in human tumours. In the mice tumour study, we detected low-level upregulation in the immune co expression network in HS and CA diet groups (Figure 5G) that had a higher tumour burden. Conversely, the immune response (mcM1) in ChR diet fed mice was higher than that in the HS and CA groups and they consistently had a lower tumour burden similar to that seen in humans in the LICA_FR dataset. We

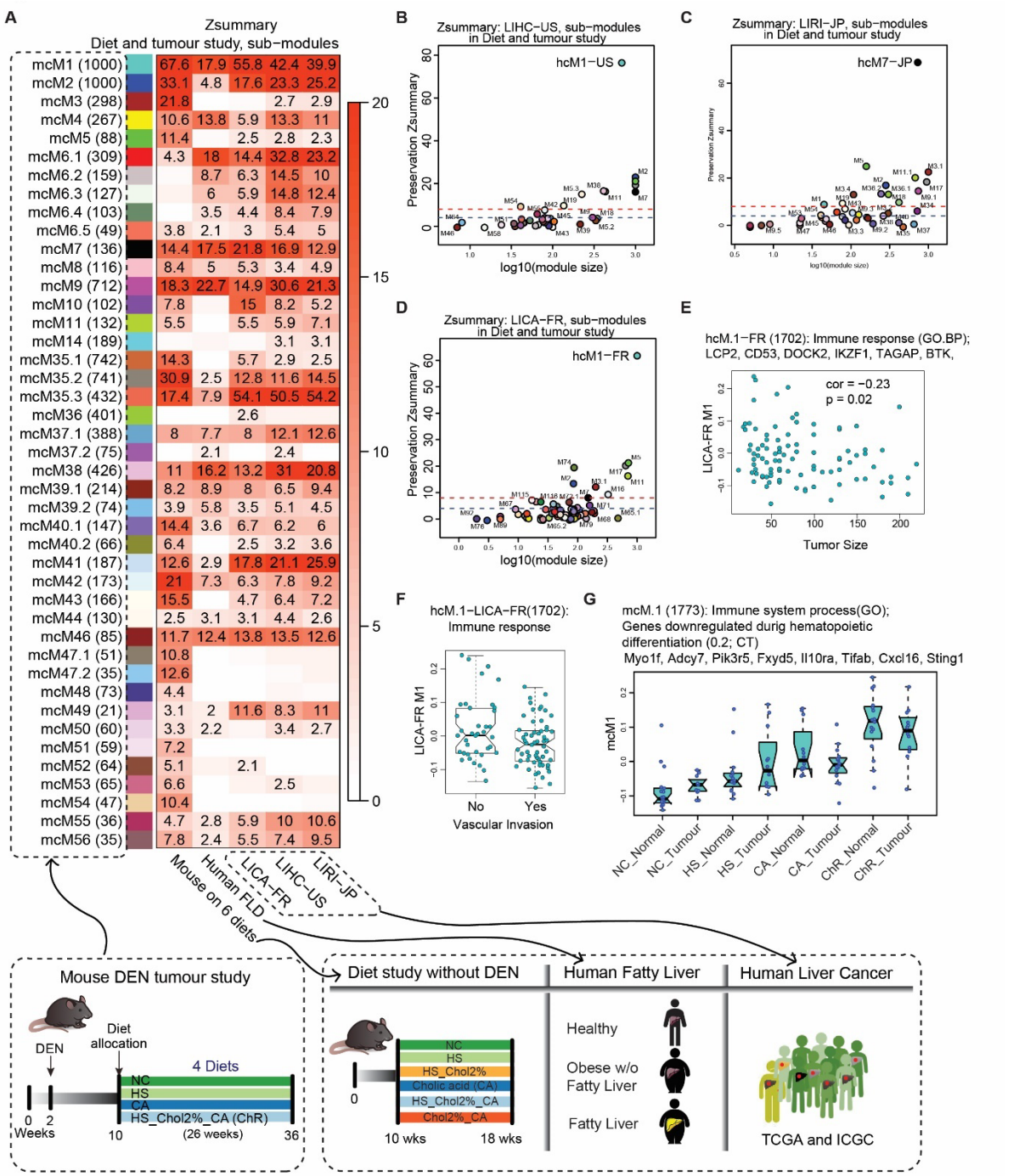

Figure 5. Reciprocal module preservation analysis demonstrates immune response is highly preserved between datasets and species. A) Heatmap representation of module preservation Zsummary statistics of diet/tumour networks constructed in DEN tumour study, in previous network analyses of a diet study without DEN, human fatty liver disease, and human liver cancer (TCGA, LIRI-JP, LICA-FR). B-D) Results of converse preservation analysis of modules identified in TCGA–US, LIR–JP, and LICA–FR datasets in mouse tumour study. The highest preserved modules are immune related modules (hcM1 in TCGA and LICA-FR, and hcM7 in LIRI-JP). E-F) Module eigengene of the immune response module hcM1 (LICA-FR) is negatively correlated with tumour size (cor= -0.23, p=0.02) and lower hcM1 level is associated with higher vascular invasion. G) Module eigengene "expression" of mouse cancer module M1 (mcM1) across the diet/condition groups. Each box represents the inter-quartile range and the thick line in the box represents the median. Whiskers extend to the most extreme point that is no further than 1.5 times the interquartile range. Blue dots represent actual eigengene expression values for each sample. The title shows the enrichment label for the module (top enrichment terms) and the top hub genes (ordered by decreasing kME).

postulated that exploring this outcome would offer novel insights.

The overlap between co-expression networks in the current study and our previous 8 week diet study without DEN is presented in Supplementary Figure 5A. As mentioned, mcM1 (enriched in immune response-related terms, hub genes: *Myo1f*, *Adcy7*, *Pik3r5*, *Fxyd5*, *Il10ra*, …, current study) has high overlap with mM1 (previous diet study without DEN). Likewise, mcM2 related to metabolism (hub genes: *Gcdh, Sucla2, Hibadh, Uqcrc2, Adh5, Coasy,* …) overlaps highly with mM2 (Metabolism). However, comparing the response of these two modules between studies indicates that the response to a ChR diet in the current study is not as distinctive as the earlier 8 weeks study without DEN (Figure 5G and 6A). This suggests that immune response networks have behaviours that change over time conceivably under the influence of diet duration and ageing. Determining the factors that contribute to the dynamic changes in immune networks over time could help to better decipher fatty liver disease progression.

The liver immune network behaviour emulates bone marrow haematopoietic stem cells responses to diets

The top enriched gene sets in mice liver cancer module mcM1 were immune system processes and genes down-regulated during haematopoietic differentiation (Figure 5G and supplementary Figure 6D). This suggests that haematopoietic precursor cells might differentiate within the liver in the mouse models. To evaluate this possibility we first analysed flow-cytometry data on bone marrow haematopoietic stem cells from the previous 8 weeks diet study without DEN (Figure 6B). This indicated a higher number of Lin-Sca1+ ckit+ (LSK) haematopoietic stem and progenitor cells (HSPCs) in diets rich in cholesterol (Figure 6B). Indeed, the number of these cells mirrored the degree of the immune response within the liver (Figure 6A-B).

To explore whether the bone marrow stem cell response similarly reflects the liver immune response in the tumour study, we profiled the bone marrow haematopoietic compartment by flow cytometry. As expected, higher numbers of Lin-Sca1+ ckit+ (LSK) cells and multipotent progenitor cells (MPP2/3) were observed in ChR and CA diet groups (Figure 6C-E). The MPP2/3 phenotype is expected as haematopoietic stem cells (HSCs) can produce myeloid biased MPP2/3 cells which then give rise to all myeloid linages (Cabezas-Wallscheid et al., 2014; Pietras et al., 2015). We also noticed a higher trend of LSK cells and elevated MPP2/3 cells in the HS diet group. Overall, the degree of HSPC response in the various diet groups was commensurate with the liver immune response; the ChR diet showed the highest HSPCs response. This in turn likely impacted the liver immune response and tumour outcome.

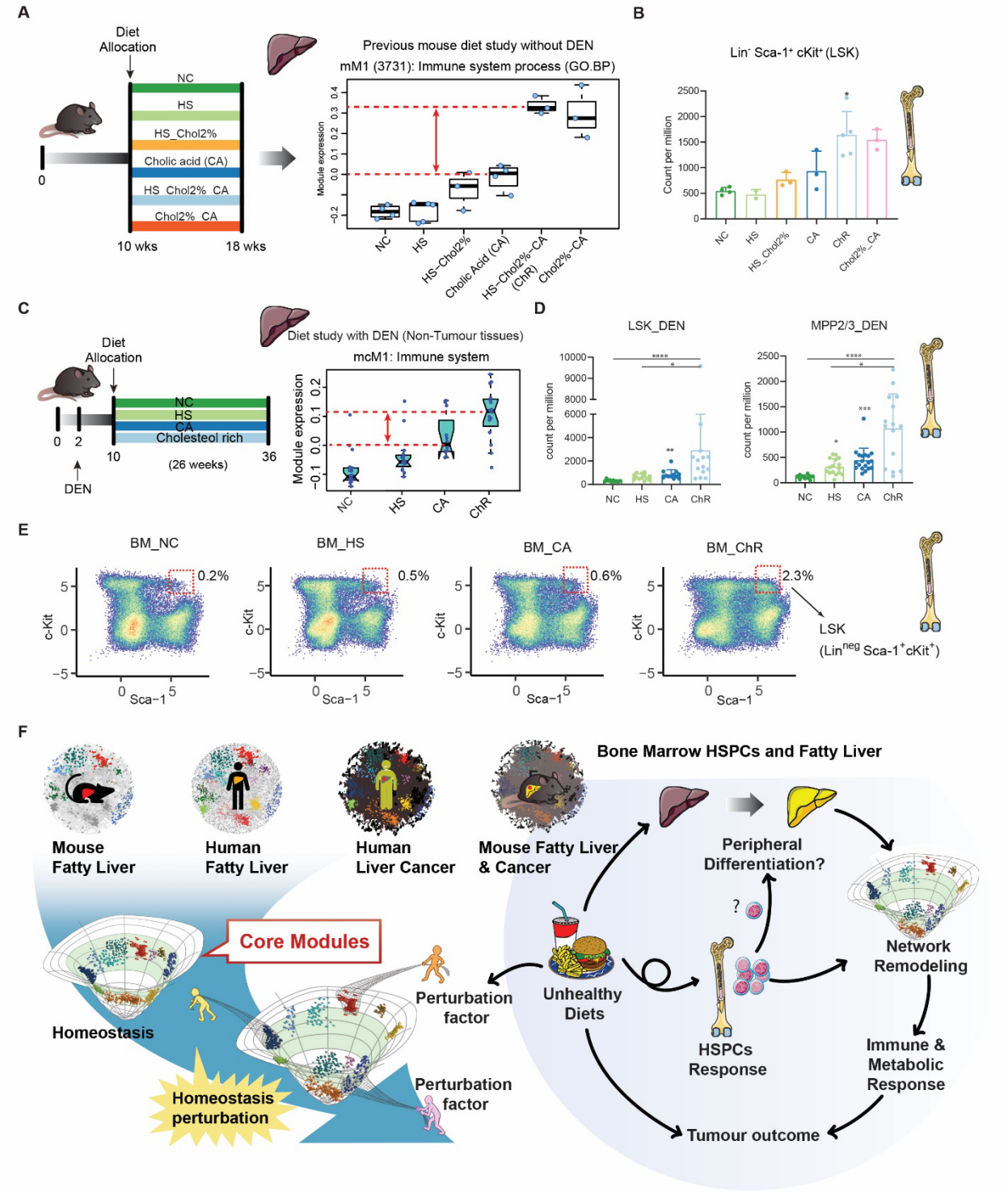

Figure 6. A coordinated response between bone marrow haematopoietic stem cells (HPSCs) and liver immune response networks under the influence of diet duration.
A) The schematics depicts a previous diet study without DEN administration. Feeding mice with diets rich in cholesterol with added cholic acid induces distinct upregulation in the liver immune response module. The module eigengene of the immune system module M1 in this network is not different between normal chow and high sucrose diet fed mice. B) BM haematopoietic stem and progenitor cells (HSPCs) response to diet mirrors liver immune responses with higher numbers of HSPCs in diets rich in cholesterol with added cholic acid (ChR & Chol2%_CA). C) Immune response network in non-tumour tissues in the mouse DEN tumour study. The immune response network is upregulated in ChR diet fed mice, although there is a less distinction between ChR and other groups. D) Higher number of Lin-Sca-1+ cKit+ (LSK) and multipotent

progenitor cell 2/3 (MMP2/3) in response to diets. HSPCs response is upregulated in HS and CA groups with distinct upregulation in ChR diet fed mice. Error bars represent mean ±1 standard deviation (SD). One-way ANOVA was used to test for significant differences in means with Dunn's (non-parametric) multiple comparison post-hoc test (Figs.6B & 6D). ****, ***, ** and * indicates a significant difference compared to normal chow, with P <0.0001, P<0.001, P<0.01 and P<0.05 respectively. E) Representative flow cytometry plots for bone marrow LSK cells. F) Schematic depicts core-liver homeostatic co-expression networks (left-previous study) and a plausible and novel mechanism linking fatty liver disease to its complications based on findings from our studies (right-current study). Our data indicates that consumption of unhealthy foods leads to homeostatic perturbations within the liver and bone marrow. Stimulation of bone marrow HSPCs and the degree of their response is linked to the extent of the liver immune response. The degree of the immune response and the severity of pro-tumorigenic risk factors determines tumour outcomes.

DISCUSION

Deconstructing the components of a biological system to its constitutive building blocks (modules) and studying their relatedness can provide insights on the function of the system *in toto* (Hartwell et al., 1999). Perturbing their function and understanding the consequent phenotypic outcomes is the next step in fathoming module behaviour to inform therapy. To do this in the context of fatty liver and HCC, we used a core-based network medicine approach and an intervention (diet) to demonstrate that core networks in liver on exposure to different diets behave as a complex and non-linear system. We show that maintaining a high immune response is associated with a reduction of liver tumour burden. Importantly, the local liver immune response reflected the bone marrow HSPC response to the various diets. A role for diet on HSPCs in fatty liver disease has not previously been reported, nor has any direct role for HSPCs in shaping the immune response to liver cancer been described. These observations highlight HSPCs as a novel accompaniment of fatty liver disease with implications for therapy. It suggests for example that strategies to target the liver *per se* might be insufficient to restore hepatic homeostasis in the context of unhealthy dietary patterns - a principal cause of fatty liver disease.

A complex interaction between tumours, the host, and environmental factors determines the strength of the anti-cancer immune response. This immune response can be conceptualised as a "cancer–immune set point", defined as the threshold that determines cancer outcomes. A set point above the threshold results in the immune system having a protective effect (Chen and Mellman, 2017). Kotliarov *et. al.*,(Kotliarov et al., 2020) utilised blood transcriptional signatures and gene co-expression modules to determine the baseline immune set-point to predict responses to vaccination or the risk of flares in systemic lupus erythematosus. Using analogous concepts, we quantitatively measured the overall effect of extrinsic factors (e.g., exposome, in our case diet) on liver gene co-expression modules and tracked the hepatic immune response. A better immune response in the ChR diet and its protective effect on cancer growth implies an above set-point immunity level; the HS and CA diets conversely were below the set point resulting in larger and more tumours. Consistently, human studies have reported that high immune responses are protective against cancer and improves overall survival (i.e. above the set-point) (Thorsson et al., 2018). In contrast, patients with an immunosuppressive tumour

microenvironment have a poor outcome (i.e. below set-point) (Thorsson et al., 2018). The bone marrow HSPC response to dietary challenge could be one of the factors regulating and maintaining the immune response set-point in our models.

Secondly, we show that extended exposure to the same diet modulates the intensity of the immune response - prolonged exposure to the ChR diet resulted in a dampened immune response module. This occurred despite a persistently high HSPC response in the marrow suggesting that while the extrinsic stimulus (diet) had not changed, local factors within the liver downregulate the liver's immune response signature. It is known that the liver is an immunotolerant organ, while Medzehitov (Medzhitov, 2021) has proposed the existence of counter-inflammatory factors (differs from anti-inflammatory) that control the responsiveness of an organ to inflammatory signals in order to restore local homeostasis. What these counter-inflammatory factors might be in the liver is uncertain but for example might include differential metabolic pathways regulation, but this would need experimental validation.

Of interest, while tumours can regulate their local immune microenvironment, they can also have distant effects including on systemic metabolism and bone marrow haematopoiesis; these effects are known as the tumour organismal environment (TOE) (Binnewies et al., 2018; Laplane et al., 2018). For example, in glioblastoma, a pro-tumorigenic role for endogenous HSPCs within the tumour microenvironment has been reported (Lu et al., 2021). Conversely, HSPCs that express C-C chemokine receptor type 2 ($CCR2^+$) reverse treatment resistance to PD-1 blockade (Flores et al., 2018). Interestingly, inciting a systemic immune response including in bone marrow and spleen is necessary for successful immunotherapy (Spitzer et al., 2017) indicating a bidirectional response which can either promote or prevent tumorigenesis. In our models, a marked HSPCs and liver immune response had protective effects.

In a recent paper, Bihari et al., reported derangement of the bone marrow haematopoietic niche and HSPC loss as a contributing factor to immune dysfunction in cirrhosis (Bihari et al., 2016). Another recent study reported similar HSPC responses to an atherogenic diet (high fat and cholesterol without cholate) in Apoe-/- and ldlr-/- mouse models, but not in C57BL6 wild type mice (3 weeks on the diet) (Heyde et al., 2021). In the present study we detected such an effect on bone marrow HSPCs in C57BL6 mice using cholesterol rich diets containing cholic acid, and with a normal chow diet with added cholic acid. This raises the possibility that exposure to excess dietary cholesterol in the context of cholic acid which is known to alter the bile acid profile (Chen et al., 2020) modulates the bone marrow HSPCs response. Consistently, it has been reported that bile acids regulate the proliferation of hematopoietic stem cells in foetal liver (Sigurdsson et al.). In the context of fatty liver disease our data indicates an important role of the haematopoietic stem cells response to liver homeostasis. This connectivity might explain the development of the extra-hepatic associations of fatty liver disease such as cardiovascular diseases.

In conclusion, we have shown that fatty liver disease secondary to various dietary challenges is just one manifestation of a multi-system disorder that includes dysregulation of the hematopoietic system. The latter communicates and contributes to the liver response which is further impacted by local counter-inflammatory factors in order to restore liver homeostasis. Conceivably, this can impact other organs such as the cardiovascular system in chronic

metabolic diseases resulting from poor dietary choices and unhealthy lifestyles. In the context of liver cancer development, we have also demonstrated that maintaining a local and above cancer–immune set point immune response in concert with a systemic bone marrow HSPC immune response reduces tumour burden. This understanding can be leveraged to improve the outcomes of patients with liver cancer.

AUTHORS' CONTRIBUTION:

Conceptualization, S.E.; Methodology, S.E., M.K.A, P.L., T.G.B; Project Administration: S.E., V.H., T.G.B; Resources: S.E., V.H., T.G.B, G.R.; Investigation, G.A.T , M.K.A., M.R-M, V.H., D.V., J.C., G.R., S.D., M.E., S.E.; Formal Analysis, P.L., D.V., C.L., B.S.G; Visualization, P.L., D.V.; Data Curation, S.E., P.L., T.G.B; Writing-Original draft, S.E.; Writing-Review & Editing, S.E., P.L., D.V., T.G.B, B.S.G, G.A.T, J.G.; Funding Acquisition; J.G., S.E.; Supervision, S.E, J.G..

ACKNOWLEDGEMENTS:

S.E. and J.G. are supported by the Robert W. Storr Bequest to the Sydney Medical Foundation, University of Sydney; National Health and Medical Research Council of Australia (NHMRC) Program (1053206) and Project grants (APP1107178, APP1108422), a Sydney West Translational Cancer Research Centre grant (IAF-2019-09, IAF-2020-02, 15/TRC/1-01), a Cancer Institute grant (ATRG 2028) and NSW Cancer Council (APP1145008). G.A.T. and D.V. are supported by a scholarship from the University of Sydney. Flow cytometry was performed in the Flow Cytometry Core Facilities that are supported by the Westmead Research Hub, the Westmead Institute for Medical Research, the Cancer Institute New South Wales, the National Health and Medical Research Council and the Ian Potter Foundation.

METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contacts (Jacob.george@sydeny.edu.au; saeed.esmaili@sydney.edu.au).

- **Source data statements**
  - Mice liver RNA-seq source data will be deposited at NCBI GEO dataset and will be publicly available
- **Code statement**
  - WGCNA original codes are publicly available at GitHub (https://github.com/plangfelder/Core-liver-homeostatic-networks).
- Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All procedures were approved by the Western Sydney Local Health District Animal Ethics Committee and conducted in accordance with Animal Experimentation guidelines of the National Health and Medical Research Council (NHMRC) of Australia. Male C57BL/6 mice were obtained from Animal Resources Centre (Perth, Australia) and used for establishing breeding colonies in Westmead Research holding facility and their litters subsequently were used in diet studies. Mice were exposed to a 12-hr light/dark cycle in pathogen free conditions with free access to food and water. Male C57BL6 mice were injected intraperitoneally with 25 mg/kg body weight DEN (Sigma-Aldrich, Munich, Germany) at 14 days of age. These mice were then given either sucrose rich diet (HS), a cholesterol rich (ChR) diet containing high sucrose with 2% cholesterol and 0.5% cholic acid (HS_Chol2%_CA), or a diet containing 0.5% cholic acid (CA) (Specialty Feed Service, Glen Forest, Australia) starting at 10 weeks of age for 26 weeks (n=20-24 mice per group). The control mice group were fed the NC (Specialty Feed Service, Glen Forest, Australia) throughout the experiment. The mice were harvested at the age of 36 weeks. At the time of harvest, mice were anesthetized with i.p. ketamine (100 mg/kg)/xylazine (10 mg/kg) injection after a 4-hr fasting period. Blood was collected by cardiac puncture. Liver was harvested, a small portion of normal liver and resectable tumours were rapidly snap frozen in liquid nitrogen and stored at -80 ºC. A thin slice of liver tissue was formalin fixed for histology examination.

Liver histology and analysis

Liver tissues were fixed in 10% neutral-buffered formalin for 24 hrs and paraffin-embedded tissue sections were stained with hematoxylin and eosin (H&E). The tissues had undergone whole tissue scanning (Hamamatsu Photonics, NanoZoomer, C9600-02). The samples were blindly assessed by DM according to the criteria for grading and staging of murine fatty liver(Kleiner et al., 2005; Liang et al., 2014) and liver tumours (Healy et al., 2015).

Glucose tolerance test

The glucose tolerance test (GTT) was performed to evaluate mice capacity in processing and storing glucose. Fasting blood glucose (FBS) (after a 3-hour fasting period) was measured by taking the blood from the tail tip using an Accu-Chek glucometer (Roche Diagnostics). Mice then received an intraperitoneal (ip) injection of glucose (1.5 mg/g), and blood glucose readings were taken at 15-, 30-, 45-, 60- and 120-mins post injection.

Biochemical assays

Serum ALT and AST were measured by automated techniques within the Department of Clinical Chemistry, Westmead Hospital. Serum lipid panel was measured using Cobas b 101 POC system and its lipid disc (both from Roche Diagnostics).

Flow cytometry analysis

Raw FCS files were compensated, arcsinh transformed (a=0, b=1/150, c=0) cleaned (margin events, debris, singlets, live) and cell populations of interest (Lin-,Sca-1+,c-Kit+ [LSK], Lin-,Sca-1-,c-Kit+ [LKS-] and residual Lin-) were gated using functions provided by ggcyto_1.18.0, openCyto_2.2.0, flowStats_4.2.0 and flowCore_2.2.0 (Ellis et al.; Finak et al., 2014; Hahne et al., 2009; Van et al., 2018). CATALYST_1.14.0 (Nowicka et al., 2017) provided wrapper functions for FlowSOM and ConsensusClusterPlus. Pre-gated LSK, LKS- and residual Lin- cells were clustered based on cell marker expression (fluorescence intensity). Similar clusters were manually merged into cell subtypes of interest as defined here (Pilzecker et al., 2017). Plots were generated using ggplot2_3.3.3. The glm function provided by glmm_1.4.2 was used to fit linear models (family = gaussian) where cells/million live singlets of studied populations and diet were the response and independent variables, respectively. Pairwise adjusted Pvalues were generated using Tukey's honest significance test using the glht function provided by multcomp_1.4-17. The analyses were performed in R 4.0.3 global environment using RStudio IDE 1.2.5033.

RNA_Seq

Total tumour and non-tumour liver RNA were extracted using RNAeasy (Qiagen). RNA purity and integrity were confirmed using Tapestation. Libraries were prepared from 100 ng total RNA using Illumina Truseq stranded mRNA library prep kit. The library average fragment length and standard deviation (SD) was verified using Tapestation before 100bp-SE sequencing on the Illumina NovaSeq 6000, yielding a median of 20 M reads per sample in mouse tumour study. Sequencing adapters and low-quality reads were trimmed using the 'Maximum information quality filtering' algorithm implemented in Trimmomatic-0.39. A cut-off length of 40 and strictness setting 0.999 was used (Bolger et al., 2014). The cut-off for the minimum read length was set to 36bp.

Reads were mapped using Salmon-1.4.0 (Patro et al., 2017) . against the GENCODE vM25 mouse transcriptome with GRCm38 mouse genome assembly as a decoy. The Salmon mapping algorithm was used with both random hexamer priming and GC bias correction.

The R package tximeta-1.8.2 was used to generate raw count matrices from Salmon quant.sf output files. Reads were summarised at the gene level and raw count matrices exported as csv for downstream processing.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of phenotypic data

Bar and XY graphs presenting mouse phenotypic data (Figure 1 and supplementary Figure 1) were generated in GraphPad prism version 8.4.3. Error bars represent ±1 standard deviation (SD) from the mean. The data presented are the pool of two independent experiments.

One-way ANOVA with Dunn's (non-parametric) multiple comparison post-hoc, test was used was to test for significant differences in means. Two-ways ANOVA with Tukey's test was used in Fig1B, 1H, and Supplementary Fig.1A. Statistical significance is presented as $P<0.05$ (*), $P<0.01$ (**), $P<0.001$ (***), or $P<0.0001$ (****).

WGCNA, enrichment, and preservation analysis

RNA_Seq data pre-processing

We retained only genes with at least 0.5 count per million reads in at least the number of samples in the smallest group (i.e., 12 samples). The rationale is to only include genes that are likely to be expressed in at least 1 diet/condition group.

To identify outlier samples, we used a modified version of the sample network methodology originally was described by Oldham et.al. (Oldham et al., 2012). Specifically, to quantify inter-sample connectivity, we first transformed the raw counts using variance stabilization (R function varianceStabilizingTransformation) and then used Euclidean inter-sample distance based on the scaled profiles of the 8000 genes with highest mean expression. The intersample connectivities *k* were transformed to Z scores using robust standardization,

$$Z_a = \frac{k_a - median(k)}{1.4826 \times MAD(k)},$$

where index *a* labels samples, MAD is the median absolute deviation, a robust analog of standard deviation, and the constant 1.4826 ensures asymptotic consistency (approximate equality of MAD and standard deviation for large, normally distributed samples). Finally, samples with $Z_a < -6$ were removed. This procedure resulted in the removal of 2 normal (CA diet) and 1 tumour (HS diet) samples.

To make DE testing and WGCNA analysis robust against potential outlier measurements (counts) that may remain even after outlier sample removal, we calculated individual observation weights designed to downweigh potential outliers as follows. First, separately for each gene, Tukey bi-square-like weights λ (Wilcox, 2012) were calculated for each (variance-stabilized) observation x_a (index *a* labels samples) as

$$\lambda_a = (1 - {u_a}^2)^2,$$

where

$$u_a = min\{1, |x_a - median(x)|/(9\,MMAD(x))\}.$$

The median was calculated separately for each gene across all samples. MMAD stands for modified MAD, calculated as follows. For each gene, we first set MMAD = MAD. The following conditions were then checked separately for each gene: (1) 10th percentile of the weights λ is at least 0.1 (that is, the proportion of observations with weights <0.1 is less than 10%) (Langfelder and Horvath, 2012) and (2) for each individual group (diet), 40th percentile of the weights λ is at least 0.9 (that is, at least 40% of the observation have a high coefficient of at least 0.9). If both conditions are met, MMAD = MAD. If either condition is not met, MMAD equals the lowest value for which both conditions are met. The rationale is to exclude outliers but ensure that the number of outliers is not too large either overall or in each genotype group. This approach has previously been used in (Esmaili et al., 2021)·(Lee et al., 2018).

Differential expression analysis

DE testing was carried out in R using package DESeq2 (Love et al., 2014) version 1.28.1, with each diet/group contrast as a variable of interest. We disabled replacement of outlier measurements in DESeq2 (because outliers are suppressed using weights) as well as independent filtering of the results (because low-expressed genes were filtered out in the preprocessing step.

Enrichment of sets of DE genes

We evaluated enrichment of significantly DE genes in the following collections of literature sets:

1. Gene Ontology (GO)(Blake et al., 2015);
2. Molecular Signatures Database (MSigDB) version 6.2 (Subramanian et al., 2005);
3. NCBI BioSystems gene sets(Geer et al., 2010), including KEGG, Reactome, Lipid Pathways and BioCYC;
4. Cell type markers for various immune and blood cells;
5. Liver-related sets from a collection of gene sets curated from general literature;
6. Genomic position gene sets; each set contains genes in a 5 Mb window, with two adjacent windows overlapping by 2.5 Mb;
7. Enrichr (Chen et al., 2013) 2016 ChEA library;
8. Enrichr 2015 ENCODE histone modification library;
9. Enrichr 2015 ENCODE TF ChIP-seq library;
10. Enrichr 2017 mirTarBase library (Chou et al., 2018);
11. Sets of genes DE at FDR<0.05 in our previous study of pre-clinical dietary models of fatty liver disease (Esmaili et al., 2021);
12. Sets of genes DE at FDR<0.05 in a study of human fatty liver disease(Suppli et al., 2019) .

These 12 collections contain a total of 44,483 gene sets. Down- and up-regulated gene sets were tested separately. The calculations were carried out using R package anRichment (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/GeneAnnotation/) that implements standard Fisher exact test and a multiple-testing correction across all query and refence gene sets, resulting in a stringent Bonferroni multiple testing correction.

Weighted gene co-expression network analysis

Module identification

WGCNA was carried on variance stabilized data from 3 sets of samples: all ("Diet and tumour study"), normal ("Diet and tumour study, normal samples") and tumour samples ("Diet and tumour study, tumour samples"). Briefly, WGCNA (Langfelder and Horvath, 2008) starts by constructing a Topological Overlap (TO) matrix (Zhang and Horvath, 2005b). To this end, we used weighted correlation with individual sample weights determined as described above and the "signed hybrid" network in which negatively correlated genes are considered unconnected. Based on scale-free topology analysis, we used the soft thresholding power $\beta=6$ in the analysis of all 3 data sets.

WGCNA uses a topological overlap-based dissimilarity as input to average-linkage hierarchical clustering that results in a dendrogram. Modules are identified as branches in the dendrogram using Dynamic Tree Cut (Langfelder et al., 2007). To improve stability of the found modules, we used random subsets of approximately 0.63 times the number of samples to create 50 sets of perturbed WGCNA module labels. These are then used as an additional input to Dynamic Tree Cut in the final module identification. We used two

different settings of Dynamic Tree Cut to produce two sets of module labels, one resulting in larger, more robust modules (called the "main" modules) and the second that produces finer split of those main modules that can be thought of as consisting of 2 or more sub-modules. Module labels for modules from all samples were chosen such that modules that strongly overlap with modules from our previous study of pre-clinical dietary models of fatty liver disease (Esmaili et al., 2021) carry the same label, if possible. Module labels in the normal and tumour subsets were chosen similarly but the all-sample modules were used as reference.

After module identification, we analysed the enrichment of genes in each module in the same reference collections of gene sets that we used for enrichment analysis of DE genes. To make it easier to keep track of the biological function of each module suggested by enrichment analysis, we created a label ("enrichment label") for each module that collects top enriched terms from each of the 12 collections. The enrichment labels are constructed only from enrichment terms whose Bonferroni-corrected p-value is below 0.05 and the overlap of the term and module covers at least 10% of the module.

Module eigengenes lead to a natural measure of similarity (membership) of all individual genes to all modules. We define a continuous ("fuzzy") measure of module membership of gene *i* in module *I* as

$$kME_i^I = \mathrm{cor}(x_i, E^I)$$

where x_i is the expression profile of gene *i* and E^I is the eigengene of module *I*. This definition is applicable to every individual network (data set). The value of module membership lies between -1 and 1. High *kME* indicate that the expression profile of gene *i* is similar to the summary profile of module *I*. Since we use signed networks here, we consider module membership near -1 low. The advantage of using correlation to quantify module membership is that the corresponding statistical significance (p-values) can be easily computed. Genes with highest module membership are called hub genes.

Association of modules with diet and condition was tested using linear models implemented in R package limma (Ritchie et al., 2015).

Module preservation between current and previous network analyses

We carried out module preservation calculations between the current 6 network analyses and our previous study of pre-clinical dietary models of fatty liver disease (Esmaili et al., 2021), in human fatty liver data (Suppli et al., 2019) and 3 human liver cancer data sets (LICA-FR, LIHC-US and LIRI-JP) from the ICGC portal (Zhang et al., 2019). We mapped genes between mouse and human Entrez identifiers using vertebrate homology information from MGI, (http://www.informatics.jax.org/homology.shtml, downloaded March 2019). Module preservation calculations (Langfelder et al., 2011) were carried out using the modulePreservation function in WGCNA with 500 permutations. We summarized the calculated preservation statistics using the Zsummary measure.

REFERENCES

Bihari, C., Anand, L., Rooge, S., Kumar, D., Saxena, P., Shubham, S., Sukriti, Trehanpati, N., Kumar, G., Pamecha, V., *et al.* (2016). Bone marrow stem cells and their niche components are adversely affected in advanced cirrhosis of the liver. Hepatology *64*, 1273-1288.

Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., *et al.* (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med *24*, 541-550.

Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., *et al.* (2015). Gene Ontology Consortium: going forward. Nucleic acids research *43*, D1049-D1056.

Blencowe, M., and Yang, X. (2021). Found in translation-core network preservation across liver diseases and species. Cell Rep Med *2*, 100347.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D.B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., *et al.* (2014). Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell *15*, 507-522.

Chen, D.S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature *541*, 321-330.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q.N., Wang, Z.C., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics *14*.

Chen, F., Esmaili, S., Rogers, G.B., Bugianesi, E., Petta, S., Marchesini, G., Bayoumi, A., Metwally, M., Azardaryany, M.K., Coulter, S., *et al.* (2020). Lean NAFLD: A Distinct Entity Shaped by Differential Metabolic Adaptation. Hepatology *71*, 1213-1227.

Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W., Huang, W.C., Sun, T.H., Tu, S.J., Lee, W.H., *et al.* (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic acids research *46*, D296-D302.

DeGregori, J. (2017). Connecting Cancer to Its Causes Requires Incorporation of Effects on Tissue Microenvironments. Cancer Res *77*, 6065-6068.

Ellis, B., HP, H.F., Le Meur, N., Gopalakrishnan, N., Spidlen, J., Jiang, M., and Finak, G. flowCore: flowCore: Basic structures for flow cytometry data (2017). URL http://bioconductor org/packages/flowCore/ R package version *1*.

Esmaili, S., Langfelder, P., Belgard, T.G., Vitale, D., Azardaryany, M.K., Alipour Talesh, G., Ramezani-Moghadam, M., Ho, V., Dvorkin, D., Dervish, S., *et al.* (2021). Core liver homeostatic co-expression networks are preserved but respond to perturbations in an organism- and disease-specific manner. Cell Syst *12*, 432-445 e437.

Finak, G., Frelinger, J., Jiang, W., Newell, E.W., Ramey, J., Davis, M.M., Kalams, S.A., De Rosa, S.C., and Gottardo, R. (2014). OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. PLoS Comput Biol *10*, e1003806.

Finn, R.S., Qin, S., Ikeda, M., Galle, P.R., Ducreux, M., Kim, T.Y., Kudo, M., Breder, V., Merle, P., Kaseb, A.O., *et al.* (2020). Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med *382*, 1894-1905.

Flores, C.T., Wildes, T.J., Drake, J.A., Moore, G.L., Dean, B.D., Abraham, R.S., and Mitchell, D.A. (2018). Lin(-)CCR2(+) hematopoietic stem and progenitor cells overcome resistance to PD-1 blockade. Nat Commun *9*, 4313.

Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S.H. (2010). The NCBI BioSystems database. Nucleic acids research *38*, D492-496.

Hahne, F., Gopalakrishnan, N., Khodabakhshi, A.H., Wong, C., and Lee, K. (2009). flowStats: Statistical methods for the analysis of flow cytometry data. R Packag version *3*.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature *402*, C47-52.

Healy, M.E., Chow, J.D., Byrne, F.L., Breen, D.S., Leitinger, N., Li, C., Lackner, C., Caldwell, S.H., and Hoehn, K.L. (2015). Dietary effects on liver tumor burden in mice treated with the hepatocellular carcinogen diethylnitrosamine. J Hepatol *62*, 599-606.
Heyde, A., Rohde, D., McAlpine, C.S., Zhang, S., Hoyer, F.F., Gerold, J.M., Cheek, D., Iwamoto, Y., Schloss, M.J., Vandoorne, K.*, et al.* (2021). Increased stem cell proliferation in atherosclerosis accelerates clonal hematopoiesis. Cell *184*, 1348-1361 e1322.
Kleiner, D.E., Brunt, E.M., Van Natta, M., Behling, C., Contos, M.J., Cummings, O.W., Ferrell, L.D., Liu, Y.C., Torbenson, M.S., Unalp-Arida, A.*, et al.* (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology *41*, 1313-1321.
Kotliarov, Y., Sparks, R., Martins, A.J., Mule, M.P., Lu, Y., Goswami, M., Kardava, L., Banchereau, R., Pascual, V., Biancotto, A.*, et al.* (2020). Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. Nat Med *26*, 618-629.
Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics *9*, 559.
Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw *46*.
Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? PLoS Comput Biol *7*, e1001057.
Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut library for R. Bioinformatics *November*, btm563.
Laplane, L., Duluc, D., Larmonier, N., Pradeu, T., and Bikfalvi, A. (2018). The Multiple Layers of the Tumor Environment. Trends Cancer *4*, 802-809.
Lee, C.Y.D., Daggett, A., Gu, X., Jiang, L.L., Langfelder, P., Li, X., Wang, N., Zhao, Y., Park, C.S., Cooper, Y.*, et al.* (2018). Elevated TREM2 Gene Dosage Reprograms Microglia Responsivity and Ameliorates Pathological Phenotypes in Alzheimer's Disease Models. Neuron *97*, 1032-1048 e1035.
Liang, W., Menke, A.L., Driessen, A., Koek, G.H., Lindeman, J.H., Stoop, R., Havekes, L.M., Kleemann, R., and van den Hoek, A.M. (2014). Establishment of a general NAFLD scoring system for rodent models and comparison to human liver pathology. PLoS One *9*, e115922.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
Lu, I.N., Dobersalske, C., Rauschenbach, L., Teuber-Hanselmann, S., Steinbach, A., Ullrich, V., Prasad, S., Blau, T., Kebir, S., Siveke, J.T.*, et al.* (2021). Tumor-associated hematopoietic stem and progenitor cells positively linked to glioblastoma progression. Nat Commun *12*, 3895.
Medzhitov, R. (2021). The spectrum of inflammatory responses. Science (New York, NY) *374*, 1070-1075.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science (New York, NY) *298*, 824-827.
Mittal, S., El-Serag, H.B., Sada, Y.H., Kanwal, F., Duan, Z., Temple, S., May, S.B., Kramer, J.R., Richardson, P.A., and Davila, J.A. (2016). Hepatocellular Carcinoma in the Absence of Cirrhosis in United States Veterans is Associated With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol *14*, 124-131.e121.
Nowicka, M., Krieg, C., Weber, L.M., Hartmann, F.J., Guglietta, S., Becher, B., Levesque, M.P., and Robinson, M.D. (2017). CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. F1000Research *6*.
Oldham, M.C., Langfelder, P., and Horvath, S. (2012). Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. BMC systems biology *6*, 63.
Paradis, V., Zalinski, S., Chelbi, E., Guedj, N., Degos, F., Vilgrain, V., Bedossa, P., and Belghiti, J. (2009). Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liver fibrosis: a pathological analysis. Hepatology *49*, 851-859.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature methods *14*, 417-419.

Pfister, D., Nunez, N.G., Pinyol, R., Govaere, O., Pinter, M., Szydlowska, M., Gupta, R., Qiu, M., Deczkowska, A., Weiner, A.*, et al.* (2021). NASH limits anti-tumour surveillance in immunotherapy-treated HCC. Nature *592*, 450-456.
Pietras, E.M., Reynaud, D., Kang, Y.A., Carlin, D., Calero-Nieto, F.J., Leavitt, A.D., Stuart, J.M., Gottgens, B., and Passegue, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. Cell Stem Cell *17*, 35-46.
Pilzecker, B., Buoninfante, O.A., van den Berk, P., Lancini, C., Song, J.-Y., Citterio, E., and Jacobs, H. (2017). DNA damage tolerance in hematopoietic stem and progenitor cells in mice. Proceedings of the National Academy of Sciences *114*, E6875-E6883.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.
Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet *31*, 64-68.
Sigurdsson, V., Takei, H., Soboleva, S., Radulovic, V., Galeev, R., Siva, K., Leeb-Lundberg, L.M.F., Iida, T., Nittono, H., and Miharada, K. Bile Acids Protect Expanding Hematopoietic Stem Cells from Unfolded Protein Stress in Fetal Liver. Cell Stem Cell.
Spitzer, M.H., Carmi, Y., Reticker-Flynn, N.E., Kwek, S.S., Madhireddy, D., Martins, M.M., Gherardini, P.F., Prestwood, T.R., Chabon, J., Bendall, S.C.*, et al.* (2017). Systemic Immunity Is Required for Effective Cancer Immunotherapy. Cell *168*, 487-502.e415.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.
Suppli, M.P., Rigbolt, K.T.G., Veidal, S.S., Heeboll, S., Eriksen, P.L., Demant, M., Bagger, J.I., Nielsen, J.C., Oro, D., Thrane, S.W.*, et al.* (2019). Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. Am J Physiol Gastrointest Liver Physiol *316*, G462-G472.
Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A.*, et al.* (2018). The Immune Landscape of Cancer. Immunity *48*, 812-830 e814.
Van, P., Jiang, W., Gottardo, R., and Finak, G. (2018). ggCyto: next generation open-source visualization software for cytometry. Bioinformatics *34*, 3951-3953.
Wilcox, R.R. (2012). Introduction to robust estimation and hypothesis testing, 3rd edn (Amsterdam ; Boston: Academic Press).
Yizhak, K., Aguet, F., Kim, J., Hess, J.M., Kubler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N.J., Rosebrock, D.*, et al.* (2019). RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. Science (New York, NY) *364*.
Younossi, Z., Anstee, Q.M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J., and Bugianesi, E. (2018). Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastro Hepat *15*, 11-20.
Younossi, Z.M., Otgonsuren, M., Henry, L., Venkatesan, C., Mishra, A., Erario, M., and Hunt, S. (2015). Association of nonalcoholic fatty liver disease (NAFLD) with hepatocellular carcinoma (HCC) in the United States from 2004 to 2009. Hepatology *62*, 1723-1730.
Zhang, B., and Horvath, S. (2005a). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.
Zhang, B., and Horvath, S. (2005b). A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology *4*.
Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019). The International Cancer Genome Consortium Data Portal. Nature biotechnology *37*, 367-369.