Single-nucleus ATAC-seq elucidates major modules of gene regulation in the development of non-alcoholic fatty liver disease

Fumihiko Takeuchi

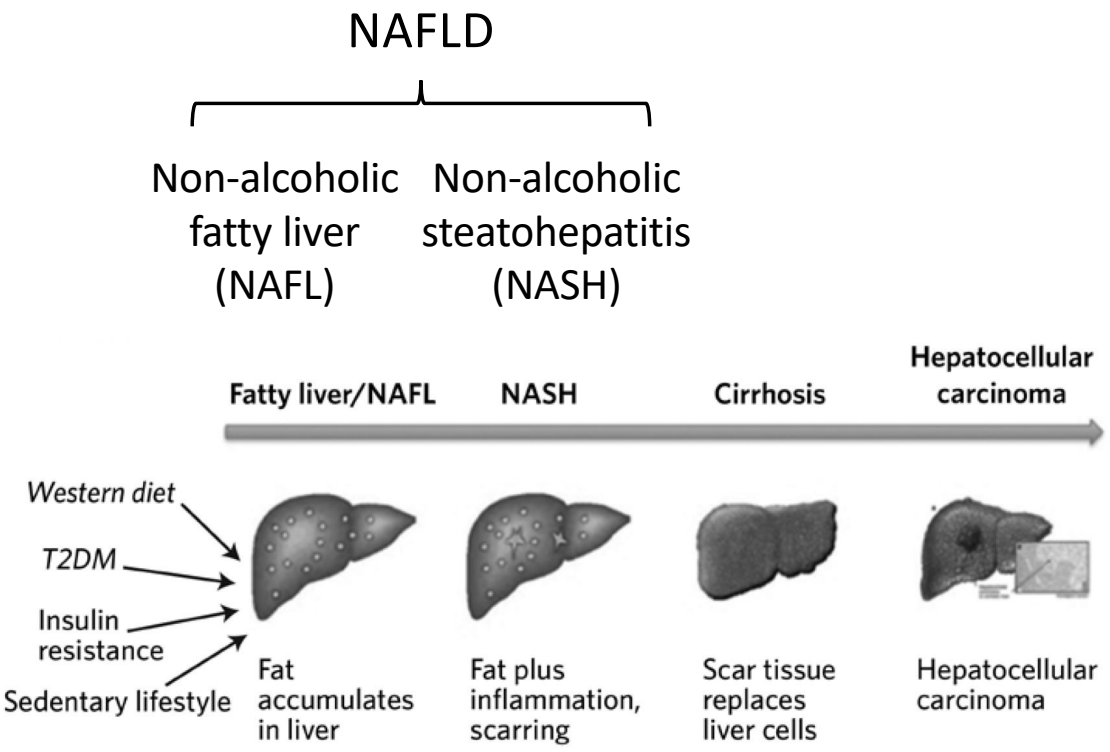
https://www.fumihiko.takeuchi.name

National Center for Global Health and Medicine (NCGM), Japan

1st Nov, 2022　Melbourne Integrative Genomics (MIG), online

Non-alcoholic Fatty Liver Disease (NAFLD)

- Global prevalence ~25%
- No approved drugs
- Proven measures
  - Weight loss by diet or exercise
  - Vitamin E
- Insufficient biomarkers

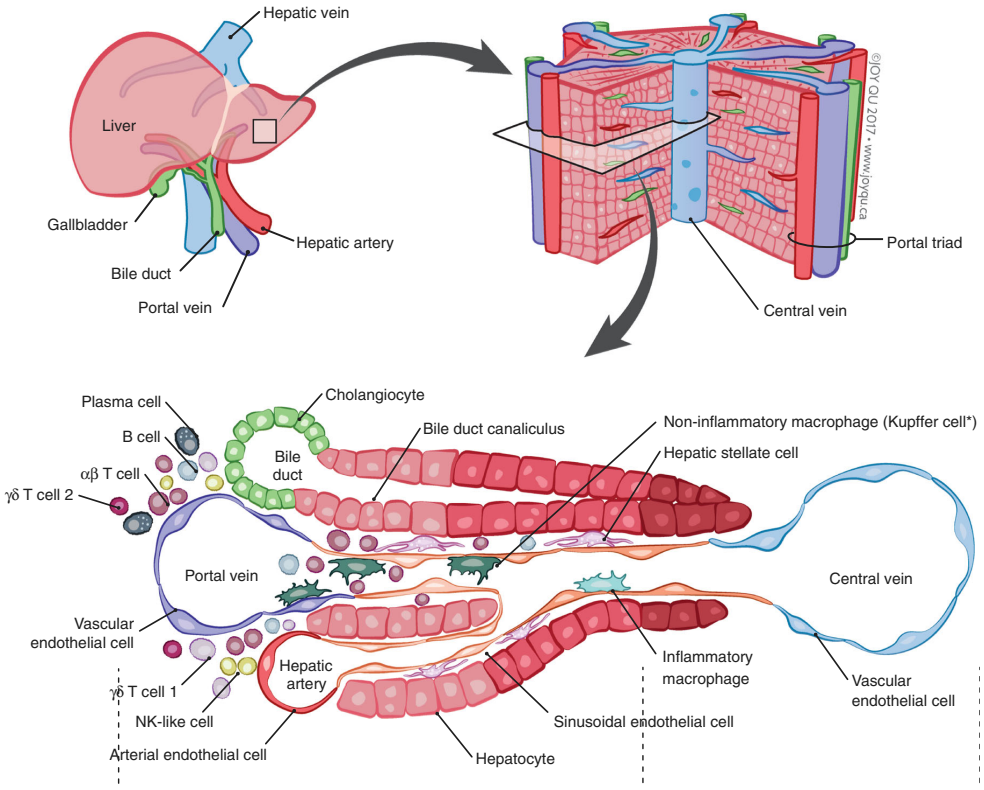


10.18043/ncm.77.3.216

Aim

- Liver tissue consists of multiple cell types
- What changes during NAFLD progression?
  1. Cell type composition
  2. Gene expression in each cell type
  3. Global gene regulation

➔ Performed single-nucleus ATAC-seq



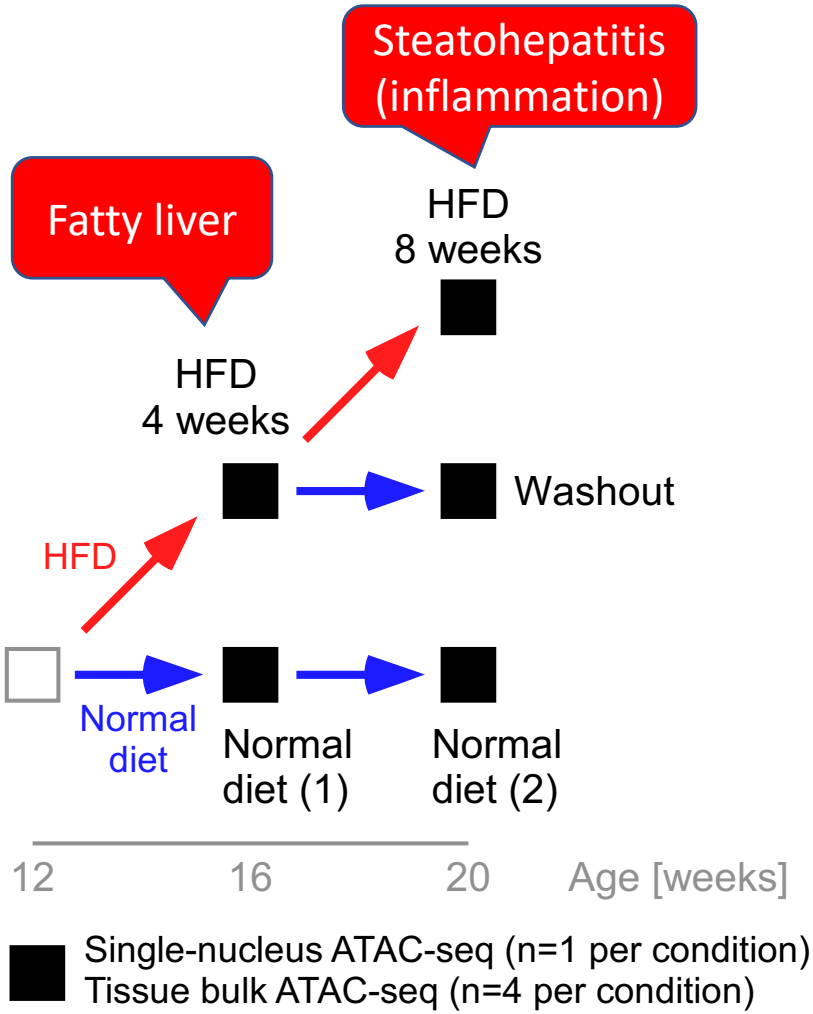
[MacParland et al. Nat Commun 9:4383]

1. Basic single-cell analysis

2. Understanding global gene regulation *in vivo*

Methods. High-fat diet model for non-alcoholic fatty liver disease

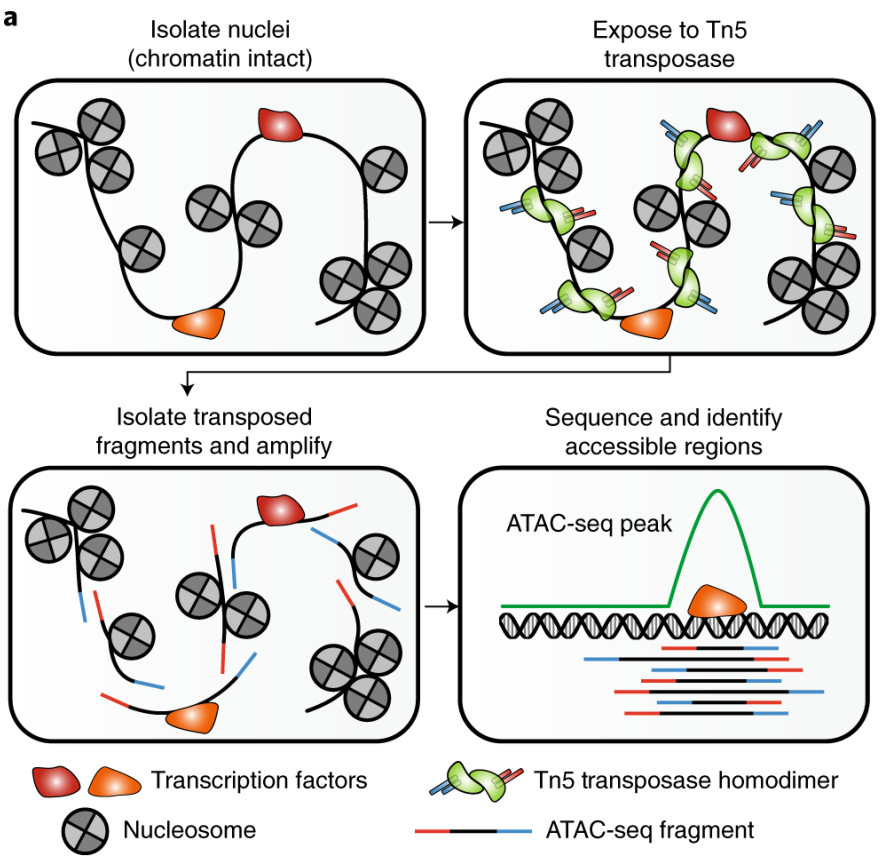
- Male Spontaneously Hypertensive Rats (SHR/Izm)
- High-fat atherogenic diet (HFD)
- ATAC-seq of livers
  - Single-nucleus ATAC-seq (10x Genomics)
  - Bulk ATAC-seq



Methods. ATAC-seq

- Detect open chromatin
  - Tn5 inserts at open chromatin
  - Insertion site is observed by NGS
- If there is nearby
  - Gene ➔ the gene is expressed
  - Transcription factor binding motif ➔ the transcription factor is binding (indirect evidence)
  - Measure both genome-wide

	ChIP-seq	Single-cell ATAC-seq
Transcription factor	1	All
Cell	Bulk aggregate	Distinguish each



10.1038/s41596-022-00692-9

Preprocessing

- For each nucleus, is Tn5 inserted at a 500bp tile (1) or not (0)

14,615 nuclei

5,230,329 tiles

Tile Matrix
0/1

- Latent semantic indexing
- Row centralization
- SVD; take top 8
- Clustering (shared nearest-neighbor graph, Leiden)

Tentative clusters of nuclei

- For each cluster, generate pseudo-bulk
- Call chromatin accessibility peaks using MACS2

Retain chromatin accessibility peaks, unified across tentative clusters.
Forget everything else.

GOAL is to obtain
- Chromatin accessibility peaks
- Putative TF binding sites, promoters

Basic data in 3 matrices

- For each nucleus, is Tn5 inserted at a peak (1) or not (0)

14,486 nuclei

599,035 chromatin accessibility peaks

Peak Matrix 0/1

- Latent semantic indexing
- Row centralization
- SVD; take top 8
- Clustering (shared nearest-neighbor graph, Leiden)

Clusters of nuclei

- For each TF, quantify its **genome-wide** binding
- Aggregate peaks having the motif of a TF
- Convert to z-score using chromVAR

14,486 nuclei

692 TFs

TF Binding Matrix

- Remove batch effect (Row centralization. SVD; subtract top 2).

Downstream analysis in specific cell type

- For each nucleus, count Tn5 insertions in gene body and vicinity
- Linnorm normalization
- SAVER imputation
- Take log2
- Quantile normalization

14,486 nuclei

17,278 genes

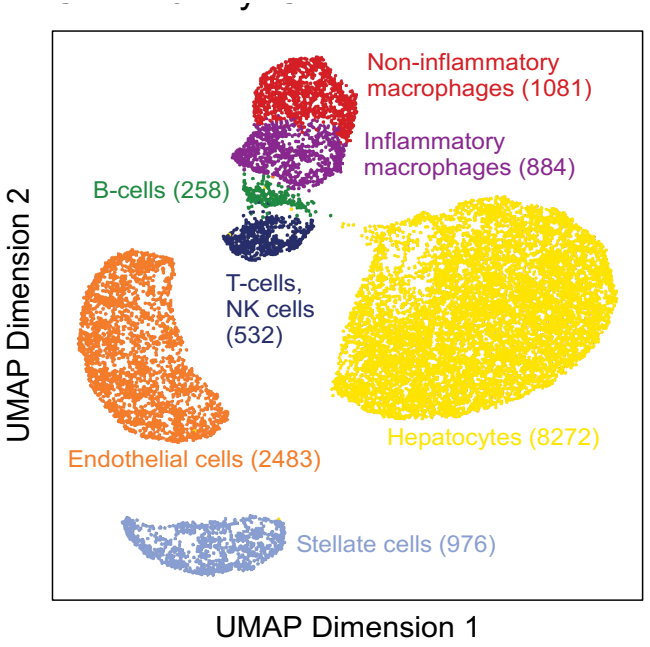
Gene Score Matrix

- Remove batch effect (Row centralization. SVD; subtract top 2).

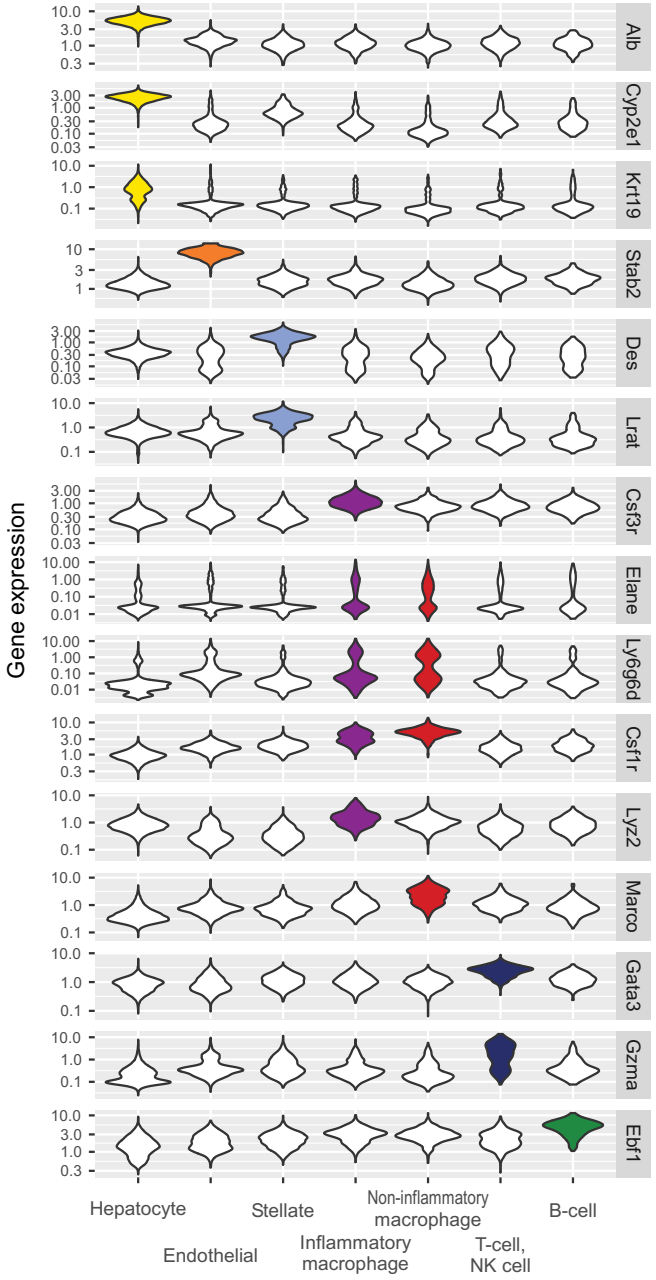
Downstream analysis in specific cell type

Observed cell types

- By similarity of chromatin opening, nuclei were grouped into 16 clusters
- Cell type assigned by marker gene expression
  - Hepatocytes
    - 7 clusters
  - Endothelial cells
    - 3 clusters
  - Stellate cells
    - 2 clusters
  - White blood cells
    - 4 clusters

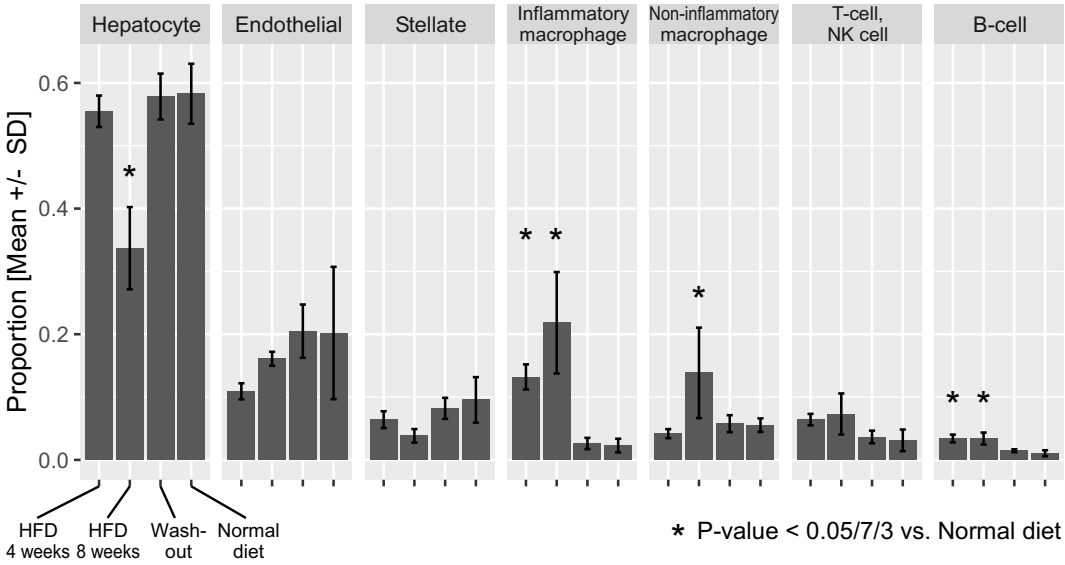


3000 nuclei per sample



Cell type composition

- Inferred cell type composition in bulk ATAC-seq samples, using snATAC-seq data as reference.
- Hepatocyte
  - Decreased after 8 weeks HFD
- Inflammatory macrophage
  - Largely increased after 4 weeks HFD
  - Further increased at 8 weeks HFD
- Non-inflammatory macrophage
  - Increased after 8 weeks HFD
- B-cell
  - Increased under HFD
- Washout didn't differ from normal diet



Differential gene expression in each cell type

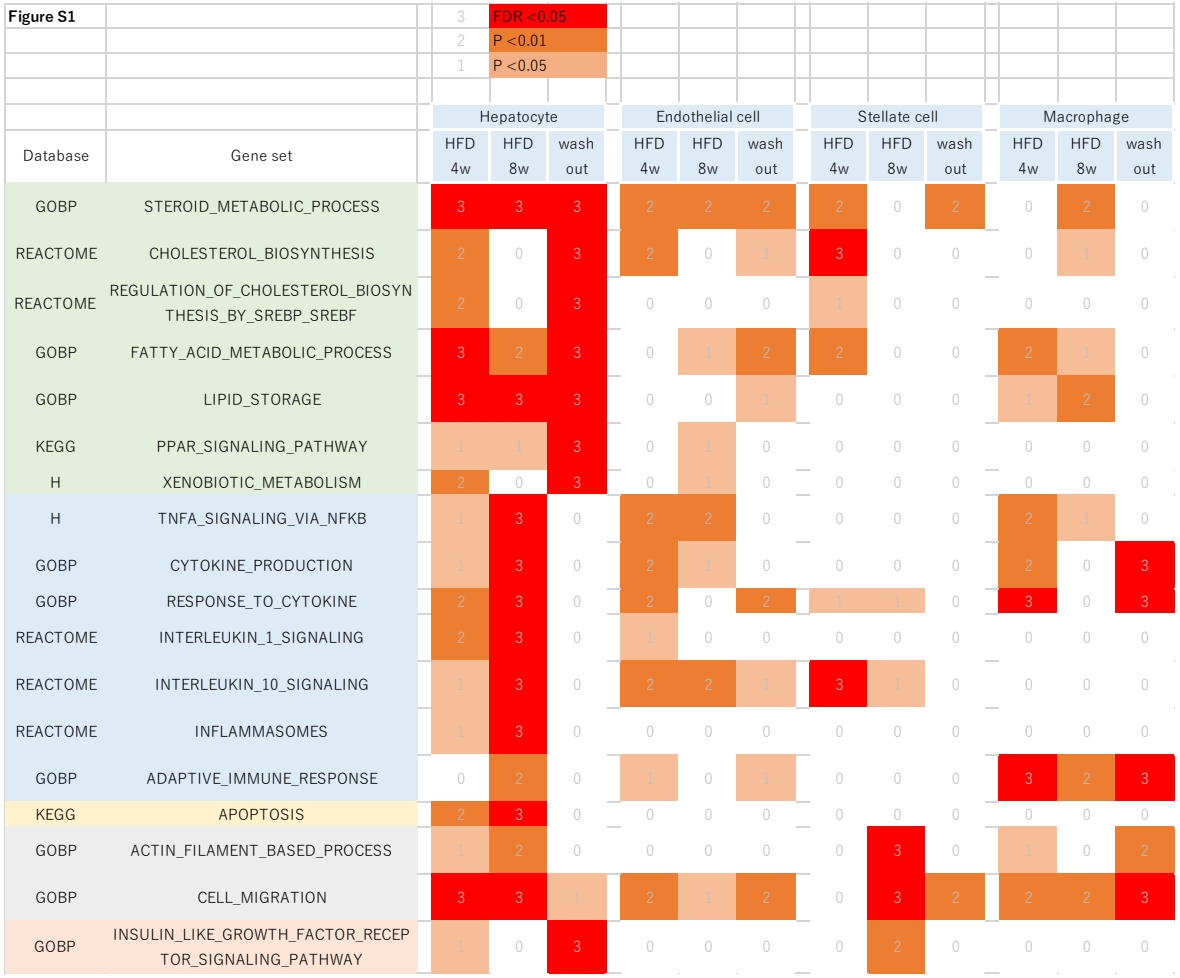


Figure S1		3	FDR <0.05											
		2	P <0.01											
		1	P <0.05											
			Hepatocyte			Endothelial cell			Stellate cell			Macrophage		
Database	Gene set		HFD 4w	HFD 8w	wash out	HFD 4w	HFD 8w	wash out	HFD 4w	HFD 8w	wash out	HFD 4w	HFD 8w	wash out
GOBP	STEROID_METABOLIC_PROCESS		3	3	3	2	2	2	2	0	2	0	2	0
REACTOME	CHOLESTEROL_BIOSYNTHESIS		2	0	3	2	0	1	3	0	0	0	1	0
REACTOME	REGULATION_OF_CHOLESTEROL_BIOSYNTHESIS_BY_SREBP_SREBF		2	0	3	0	0	0	1	0	0	0	0	0
GOBP	FATTY_ACID_METABOLIC_PROCESS		3	2	3	0	1	2	2	0	0	2	1	0
GOBP	LIPID_STORAGE		3	3	3	0	0	1	0	0	0	1	2	0
KEGG	PPAR_SIGNALING_PATHWAY		1	1	3	0	1	0	0	0	0	0	0	0
H	XENOBIOTIC_METABOLISM		2	0	3	0	1	0	0	0	0	0	0	0
H	TNFA_SIGNALING_VIA_NFKB		1	3	0	2	2	0	0	0	0	2	1	0
GOBP	CYTOKINE_PRODUCTION		1	3	0	2	1	0	0	0	0	2	0	3
GOBP	RESPONSE_TO_CYTOKINE		2	3	0	2	0	2	1	1	0	3	0	3
REACTOME	INTERLEUKIN_1_SIGNALING		2	3	0	1	0	0	0	0	0	0	0	0
REACTOME	INTERLEUKIN_10_SIGNALING		1	3	0	2	2	1	3	1	0	0	0	0
REACTOME	INFLAMMASOMES		1	3	0	0	0	0	0	0	0	0	0	0
GOBP	ADAPTIVE_IMMUNE_RESPONSE		0	2	0	1	0	1	0	0	0	3	2	3
KEGG	APOPTOSIS		2	3	0	0	0	0	0	0	0	0	0	0
GOBP	ACTIN_FILAMENT_BASED_PROCESS		1	2	0	0	0	0	0	3	0	1	0	2
GOBP	CELL_MIGRATION		3	3	1	2	1	2	0	3	2	2	2	3
GOBP	INSULIN_LIKE_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY		1	0	3	0	0	0	0	2	0	0	0	0

Gene Set Enrichment Analysis using PADOG

Compared to normal diet

- Steroid & fatty acid metabolism
  - Hepatocyte
- Inflammation
  - Macrophage: 4 weeks HFD, washout
  - Hepatocyte: 8 weeks HFD
- Apoptosis
  - Hepatocyte: 8 weeks HFD
- Actin filament
  - Stellate cell: 8 weeks HFD

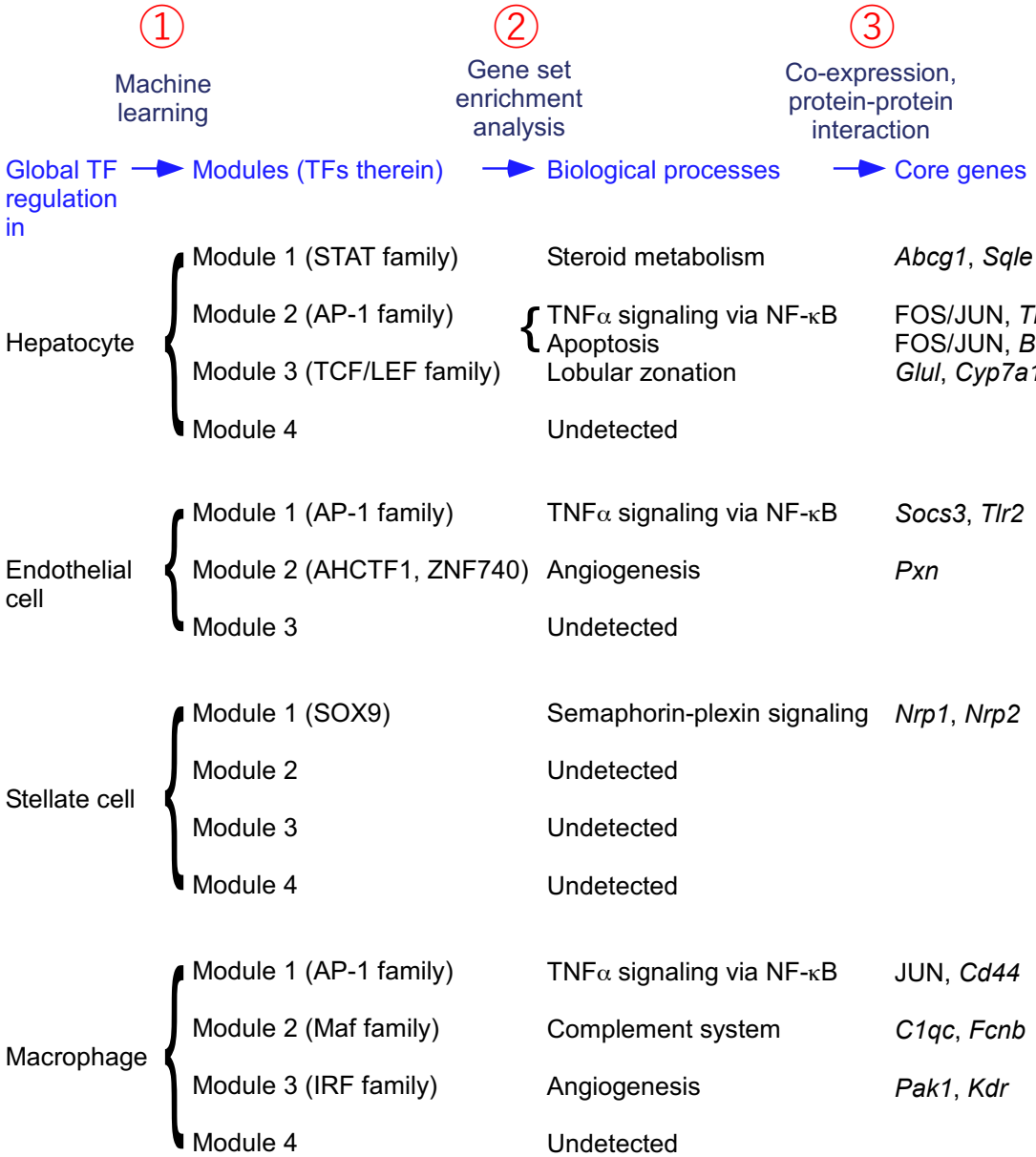
Summary (Part 1)

- By utilizing single-nucleus ATAC-seq, we could observe cell-type specific changes in the progression of NAFLD.
  - Changes in cell type composition
    - In accordance with the pathological progression, the proportion of inflammatory macrophages dramatically increased.
  - Changes in cell-type specific gene expression

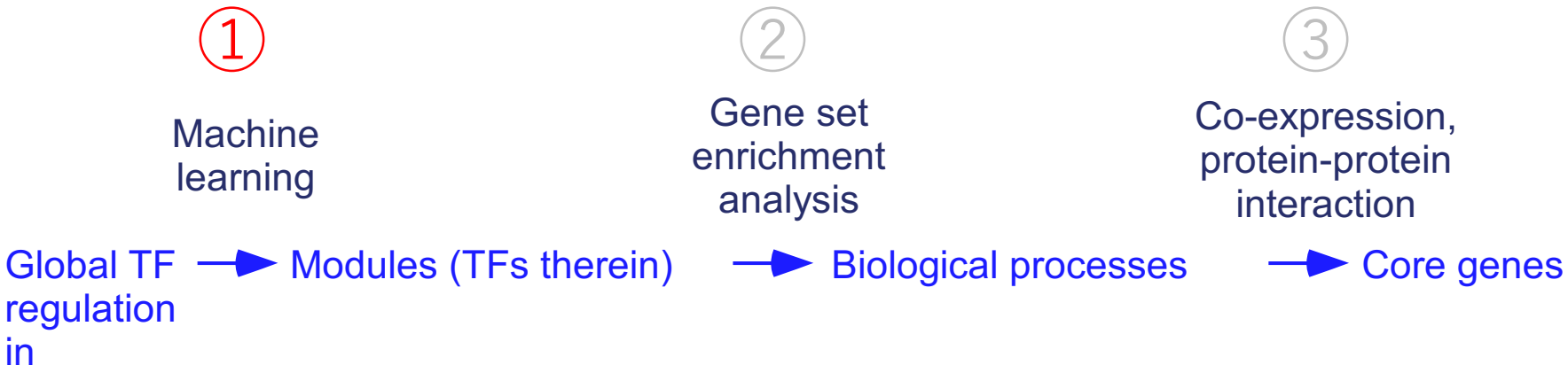
1. Basic single-cell analysis

2. Understanding global gene regulation *in vivo*

Data-driven discovery of global gene regulation

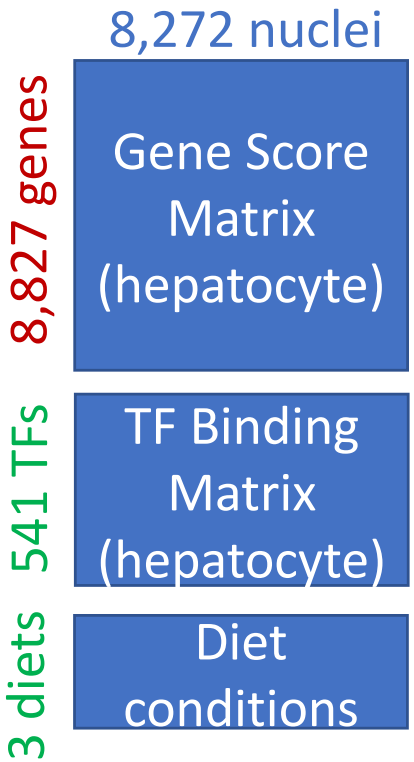


Conclusion.
Using novel statistical methods, we elucidated a global picture of *in vivo* transcription factor (TF) regulation in each cell type as a set of modules, and discovered core genes.


①
Machine learning
②
Gene set enrichment analysis
③
Co-expression, protein-protein interaction
Global TF regulation in
Modules (TFs therein)
Biological processes
Core genes

① Extract major modules of TF regulation

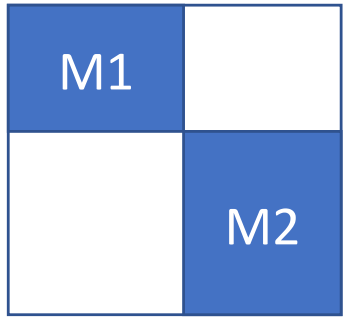
INPUT:



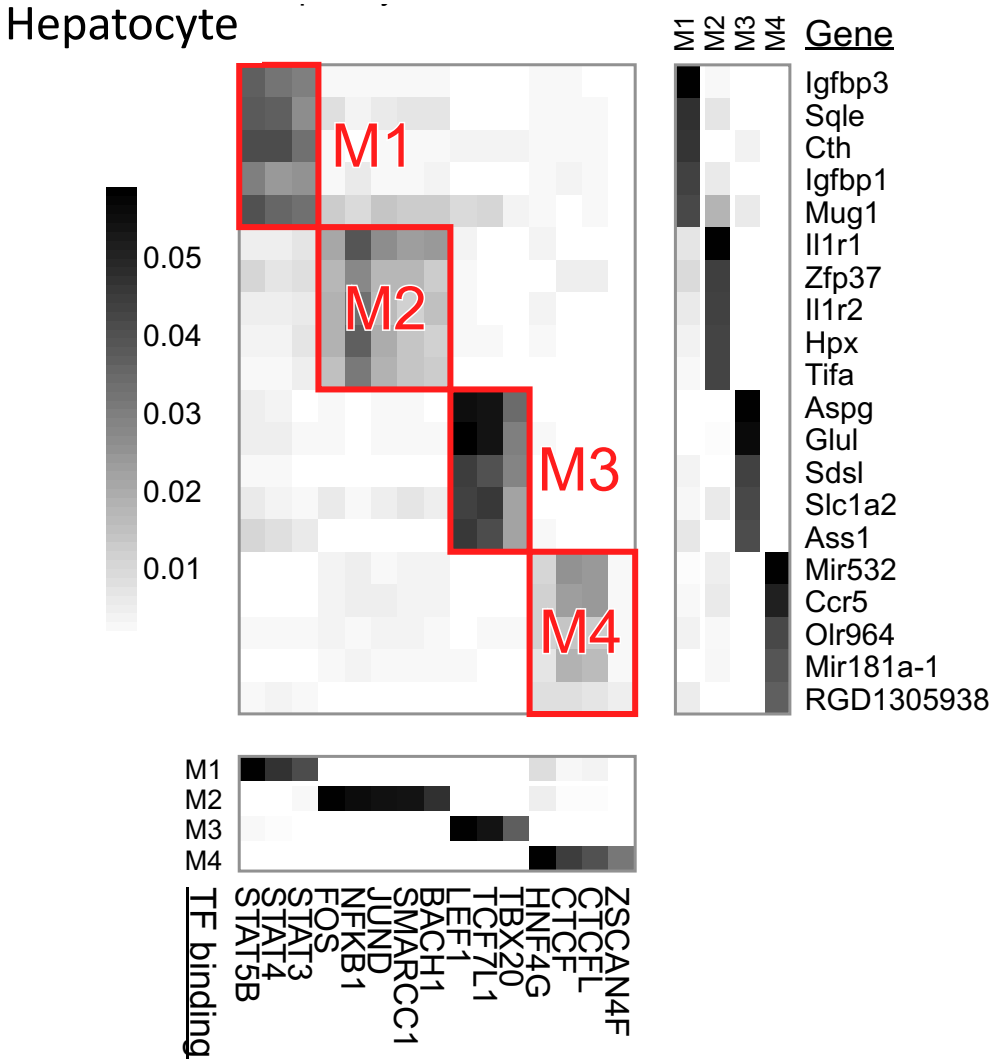
Restrict to differentially expressed TFs/genes (Foldchange>1.1, FDR<0.01)

- For each gene, compute its regulator TFs/diets
  - GENIE3 machine learning
  - Quantify as explained variance

- Extract modules
  - Nonnegative matrix factorization
  - A module is
    - a subset of TFs/genes
    - precisely, weighting of TFs/genes
  - TFs in a module regulate the genes in the same module

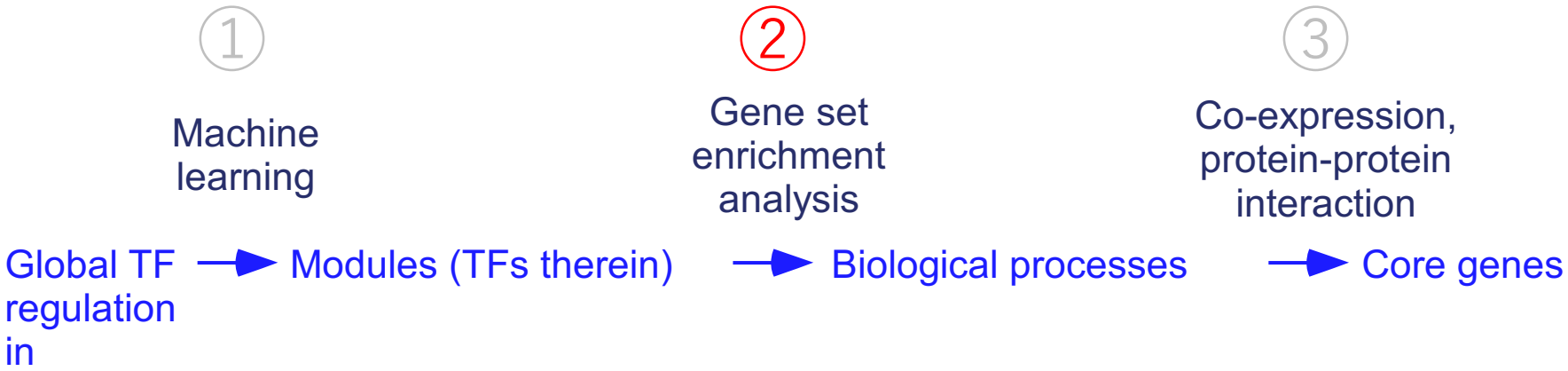


① Extract major modules of TF regulation



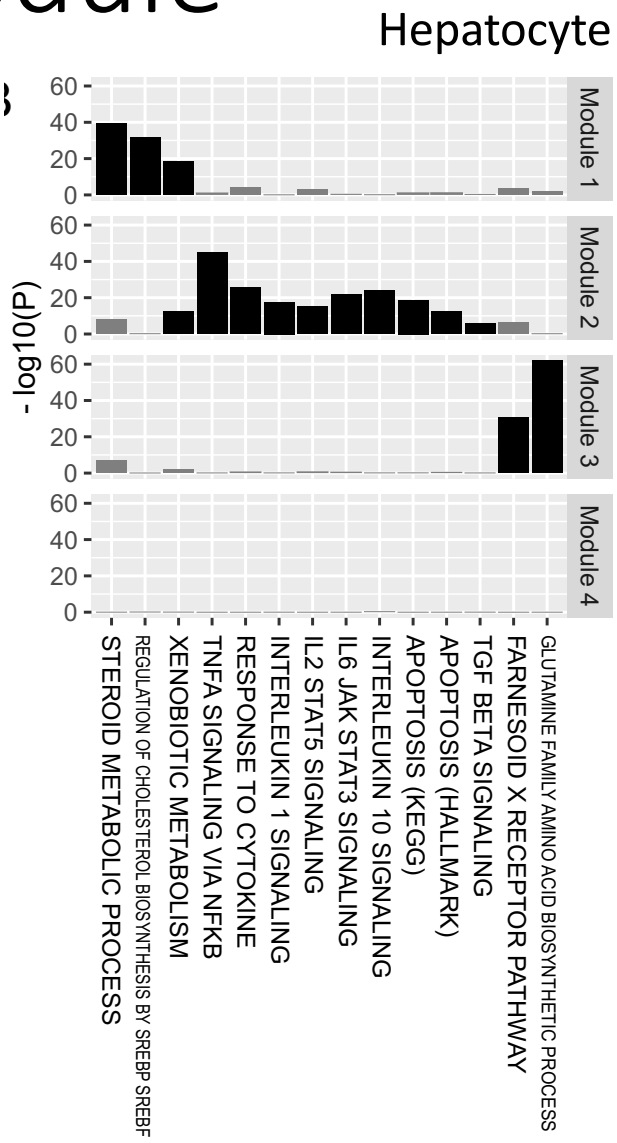
Four modules in hepatocytes

- TFs in a module regulate the genes in the same module


1
Machine learning
2
Gene set enrichment analysis
3
Co-expression, protein-protein interaction
Global TF regulation in
Modules (TFs therein)
Biological processes
Core genes

② Search biological processes characteristic to a module

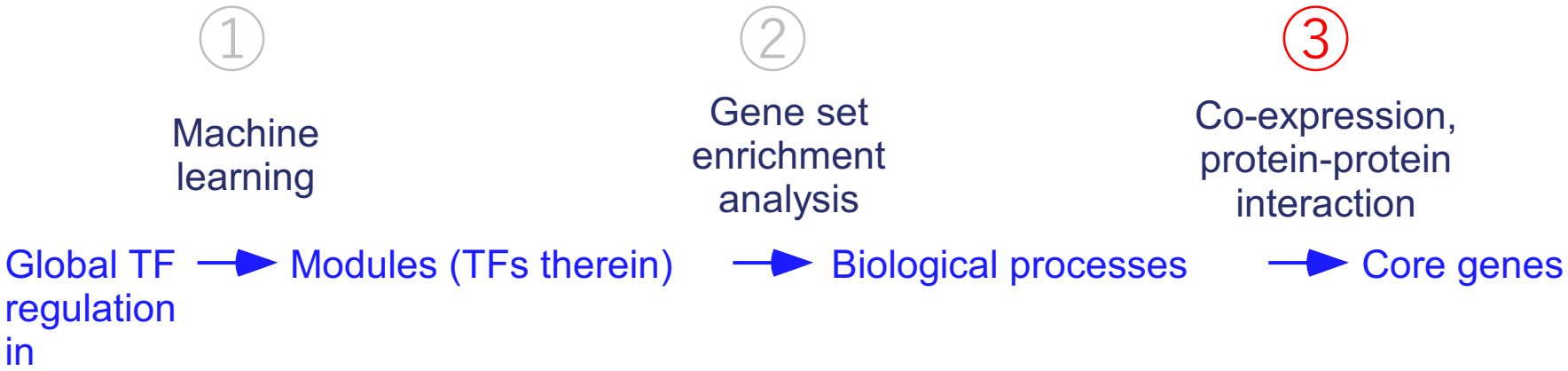
- Essentially, a module is a list of TFs and genes
- From the list of genes, find characteristic biological processes
  - Search Gene Ontology and pathway databases, using Gene Set (GS) Enrichment analysis
  - Test if heavily weighted genes were enriched in some GS
    - Weight of TFs is not used
  - Family-wise error rate of <0.05 by permuting genes



TF modules found in this study

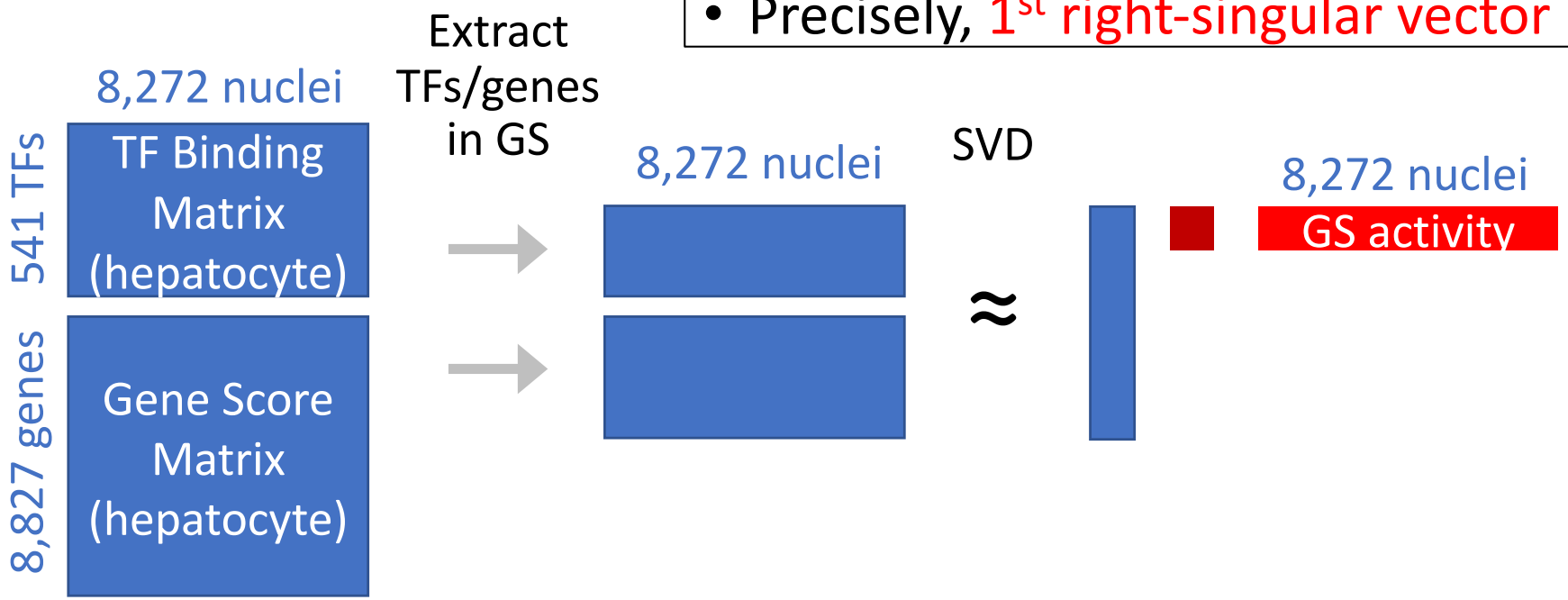
Cell type	TF	Biological process	Previous reports
Hepatocyte	STAT family	steroid metabolism	
Hepatocyte, endothelial, macrophage	AP-1 family	TNFα signaling via NF-κB	AP-1 TFs respond to cytokine stimuli (Hess et al., 2004)
Hepatocyte	TCF/LEF family	zonation in liver lobule	LEF1 TF binds to β-catenin protein and activates Wnt signaling pathway (Sun and Weis, 2011)
Endothelial	AHCTF1, ZNF740	angiogenesis	ZNF740 activates angiogenesis in pulmonary artery endothelial cells of rats (Yu et al., 2018)
Stellate	SOX9	semaphorin-plexin signaling	
Macrophage	Maf family	complement system	In *Mafb*-deficient macrophages of mice, C1q production decreased (Tran et al., 2017)
Macrophage	IRF family	angiogenesis	IRF1 contributes to the commitment of pro-inflammatory M1 macrophages, which produce angiogenic stimulators (Chistiakov et al., 2018).

- The linkage between TF and biological process has been reported in 5 out of 7
- Good indication!


①
Machine learning
②
Gene set enrichment analysis
③
Co-expression, protein-protein interaction
Global TF regulation in
Modules (TFs therein)
Biological processes
Core genes

"Gene Set (GS) activity" of each nucleus

- The "average" of TF binding and gene expression for genes in GS
- Precisely, 1st right-singular vector



- Restrict to differentially expressed TFs/genes (Foldchange>1.1, FDR<0.01)
- Standardize each row

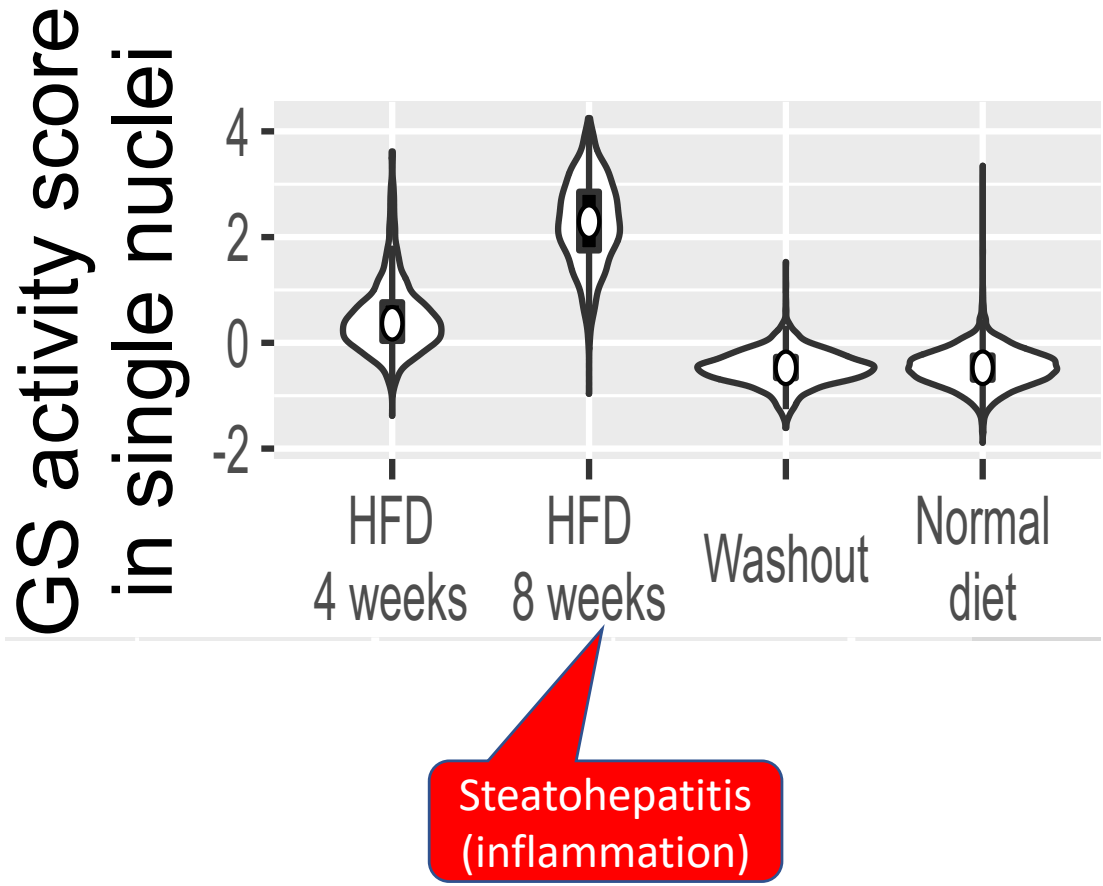
"GS activity" is meaningful only when

- TFs/genes in GS are co-expressed
- 1st singular value is significantly large
- P<0.05, permuting the extracted TFs/genes

“Gene Set (GS) activity” of each nucleus

Hepatocyte

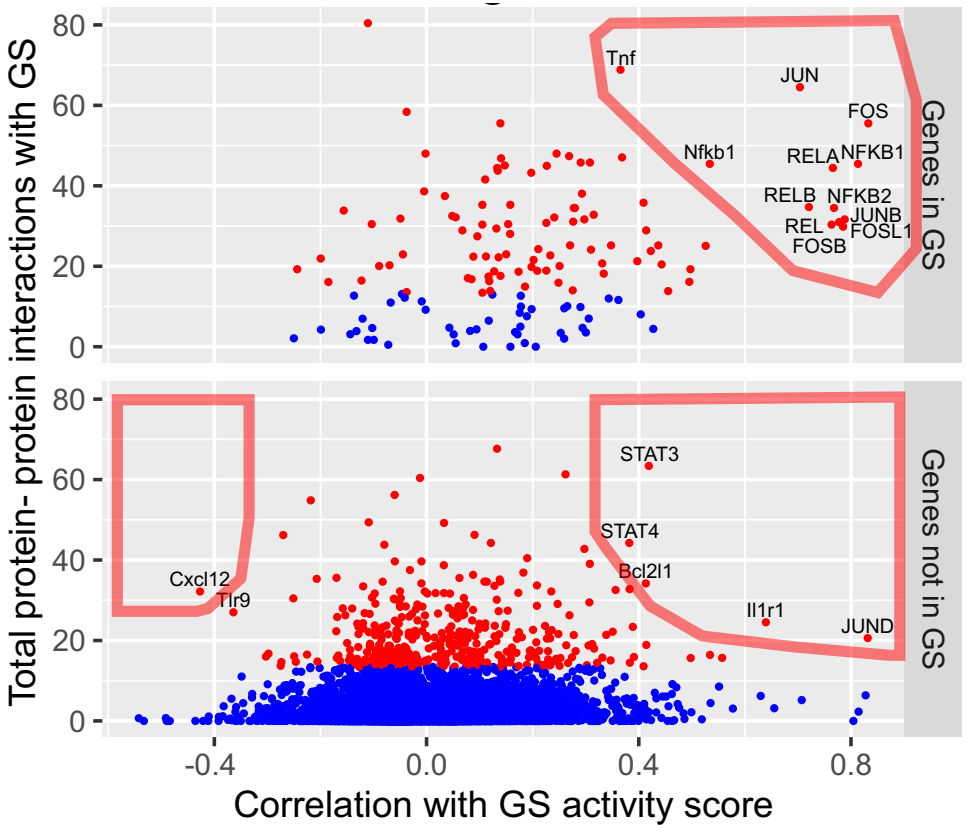
GS for TNFα signaling via NFκB (202 genes)



③ Discover core genes of a biological process

- Focus on a biological process in a cell type
  - Here, a biological process is defined as a gene set (GS)
- Core genes of a GS
  - Central in co-expression
    - Strong positive or negative correlation with “GS activity”
  - Central in protein-protein interaction
    - Many interactions with GS
    - STRING database

TNFα signaling via NFκB in hepatocytes



Core genes found in this study

Cell type	Biological process	#Core genes	Known causal or biomarker genes for NAFLD	
Hepatocyte	TNFα signaling via NF-κB	5	4	*Tnf*, *Nfkb1*, *Il1r1*, *Cxcl12* (aka *Sdf1*)
Endothelial	TNFα signaling via NF-κB	8	4	*Pecam1*, *Tlr4*, *Il15*, *Ccr5*
Macrophage	TNFα signaling via NF-κB	3	1	*Cd44*
Hepatocyte	steroid metabolism	9	3	*Scd1*, *Acox2*, *Apoa1*
Stellate	semaphorin-plexin signaling	7	3	*Nrp2*, *Nrp1*, *Sema3e*

- Large overlap with known NAFLD genes
- Suggests the biological validity of our data-driven approach

Summary (Part 2)

- We captured global gene regulation *in vivo* under high-fat diet by decomposing into modules.
- The combination of TFs and genes (and biological processes) in a module agreed with previous reports.
- Utilizing GS activity score, we searched core TFs/genes in biological processes, many of which overlapped with known NAFLD genes.

Conclusions

- By performing single-nucleus and bulk ATAC-seq, we analyzed the transition of cell type composition and cell type-specific gene expression in a rat model of NAFLD.
- Using novel statistical methods, we elucidated a global picture of *in vivo* transcription factor regulation in each cell type as a set of modules and discovered core genes for NAFLD-relevant biological processes.

Single-nucleus ATAC-seq elucidates major modules of gene regulation in the development of non-alcoholic fatty liver disease
10.1101/2022.07.12.499681
in bioRxiv