Analysis of Metabolic Dysfunction-associated Steatotic Liver Disease by Single-Cell and Spatial Transcriptomics

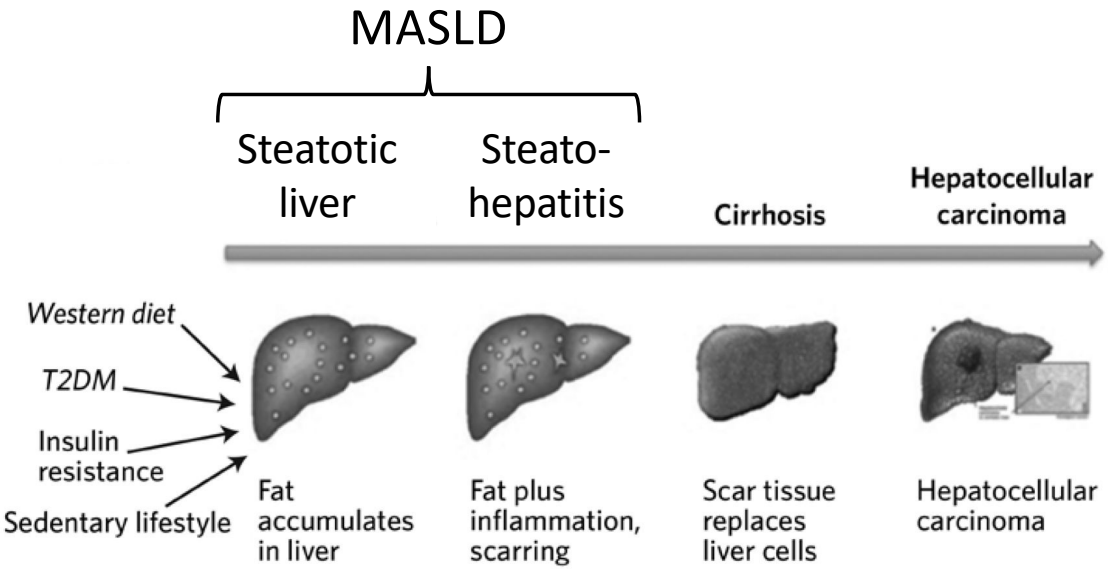
Fumihiko (Fumi) Takeuchi
Head of Bioinformatics, Systems Genomics Laboratory
5th Sep, 2023 Baker Heart and Diabetes Institute
https://www.fumihiko.takeuchi.name

Metabolic dysfunction-associated steatotic liver disease (MASLD)

Old nomenclature	New
Nonalcoholic fatty liver disease (NAFLD)	Metabolic dysfunction-associated steatotic liver disease (MASLD)
Nonalcoholic steatohepatitis (NASH)	Metabolic dysfunction-associated steatohepatitis (MASH)

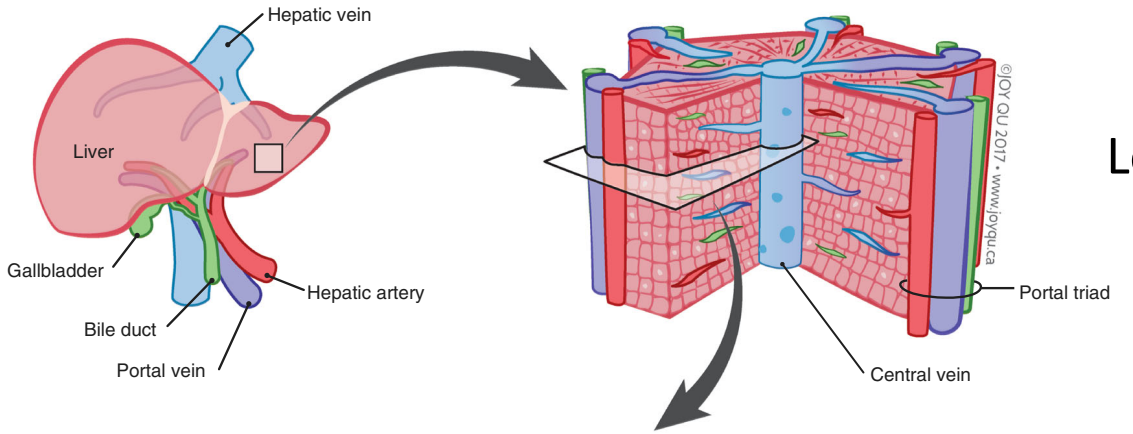
10.1016/j.jhep.2023.06.003

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



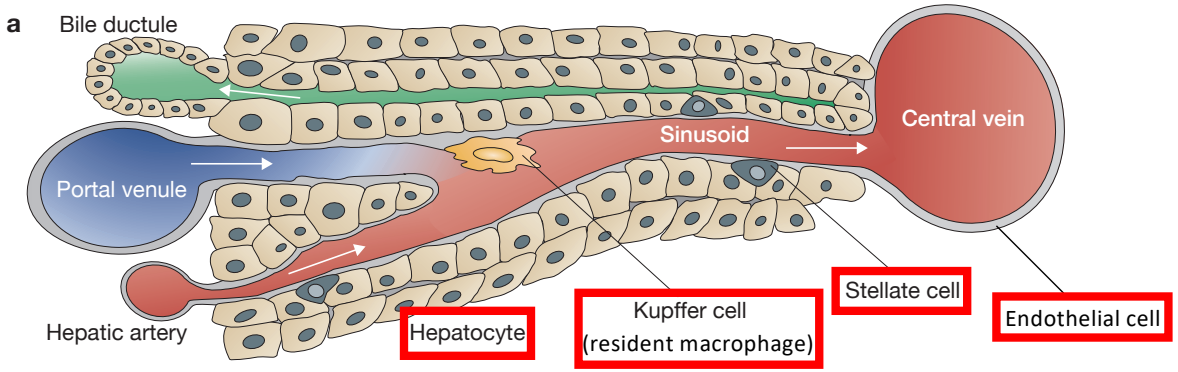
10.18043/ncm.77.3.216

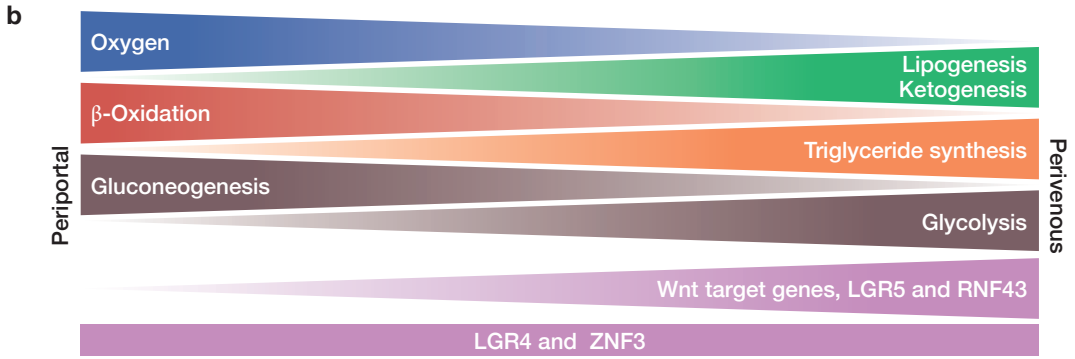
Liver, lobule, cell types and zonation



Lobule

[MacParland et al. Nat Commun 9:4383]

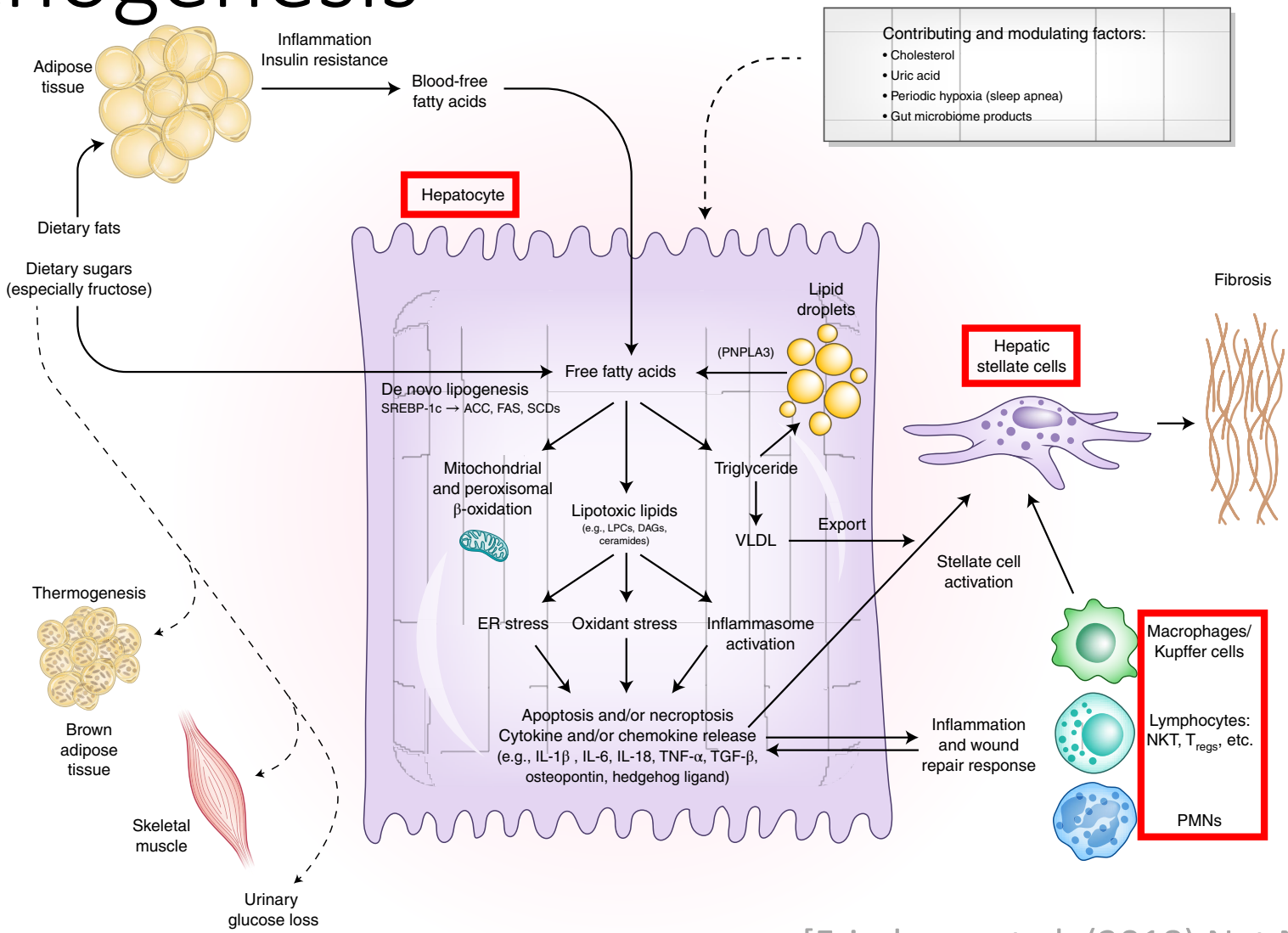




Zonation

[Birchmeier Nat Cell Biol 18:463]

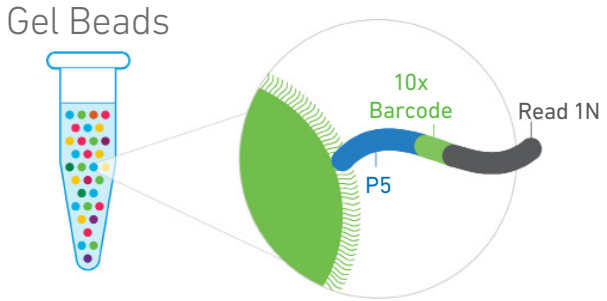
Cell types involved in MASH pathogenesis

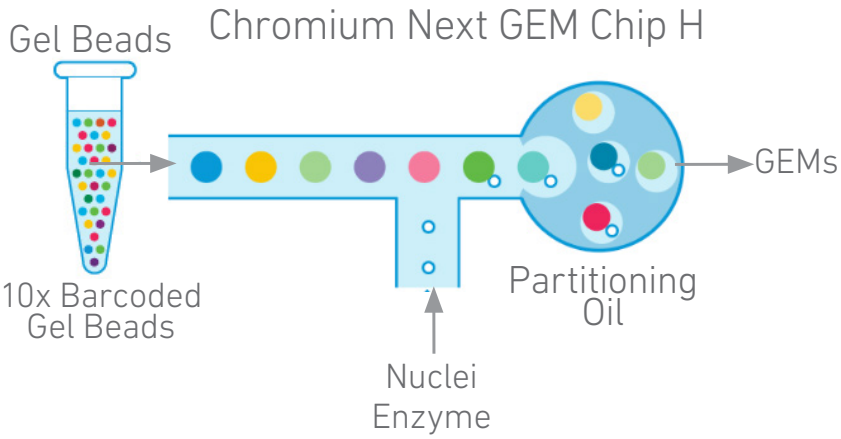


[Friedman et al. (2018) Nat Med 24:908]

Single-cell or single-nucleus assays

- Essentially, assay single cells/nuclei separately
  - e.g. in droplets
  - ~3000 nuclei per sample in this study
- Commercially available assays
  - RNA-seq
  - ATAC-seq
- We used Chromium ATAC-seq [10x Genomics]

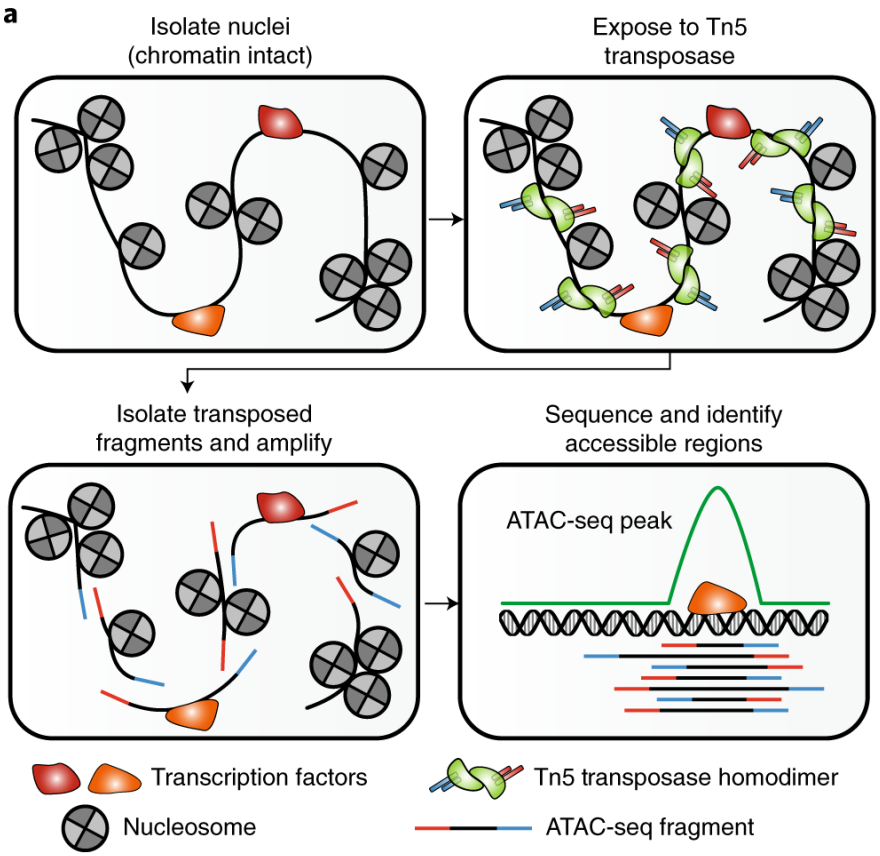




10x Genomics

ATAC-seq

- Detect open chromatin
  - Tn5 inserts at open chromatin
  - Insertion site is observed by NGS
  - Co-located insertions form a peak
- If there is nearby
  - Gene ➔ the gene is “expressed”
  - Transcription factor binding motif ➔ the transcription factor is “binding”
- Although indirect, we can quantify
  - expression of all genes
  - binding of all transcription factors (ChIP-seq/Cut&Run/Cut&Tag can only measure a few)



10.1038/s41596-022-00692-9

Published Online: 25 July, 2023 | Supp Info: http://doi.org/10.26508/lsa.202301988
Downloaded from life-science-alliance.org on 26 July, 2023

Research Article

TRANSPARENT PROCESS OPEN ACCESS Life Science Alliance

Gene-regulation modules in nonalcoholic fatty liver disease revealed by single-nucleus ATAC-seq

Fumihiko Takeuchi[1,2,3], Yi-Qiang Liang[1], Hana Shimizu-Furusawa[4], Masato Isono[1], Mia Yang Ang[1,5], Kotaro Mori[2], Taizo Mori[6], Eiji Kakazu[6], Sachiyo Yoshio[6], Norihiro Kato[1,2,5]

Part 1. Development of MASLD

- Quantity: cell type composition
- Quality: gene expression and regulation
- Single-cell assay

Part 2. Recovery from MASLD

- Distinguish recovering cells
- Single-cell + Spatial transcriptomics

Background & Aim

- Previous MASLD single-cell studies were all scRNA-seq
  - Couldn’t quantify cell type composition
  - Couldn’t measure transcription factor regulation

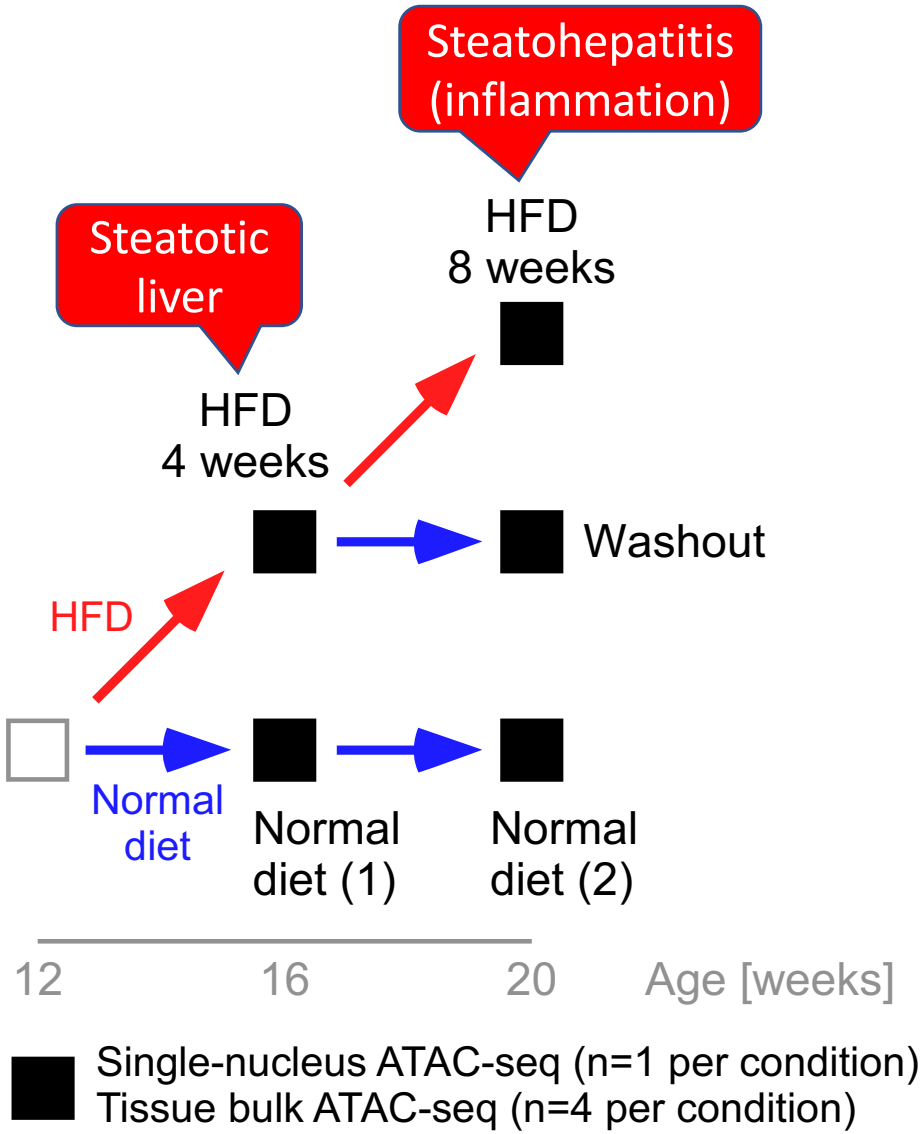
- Elucidate MASLD development
  1. Cell type composition
  2. Gene expression in each cell type
  3. Global gene regulation

➔ Performed single-nucleus and bulk tissue ATAC-seq

- Identify core genes in relevant biological processes
  - Candidates for drug targets and biomarkers

Methods. High-fat diet model for MASLD

- Male Spontaneously Hypertensive Rats (SHR/Izm)
- High-fat atherogenic diet (HFD)
  - 24% fat, 15% protein, 5% cholesterol, 2% cholic acid
- Not obese
- Not diabetic



Methods. Inferring nuclei similarity from single-nucleus ATAC-seq data

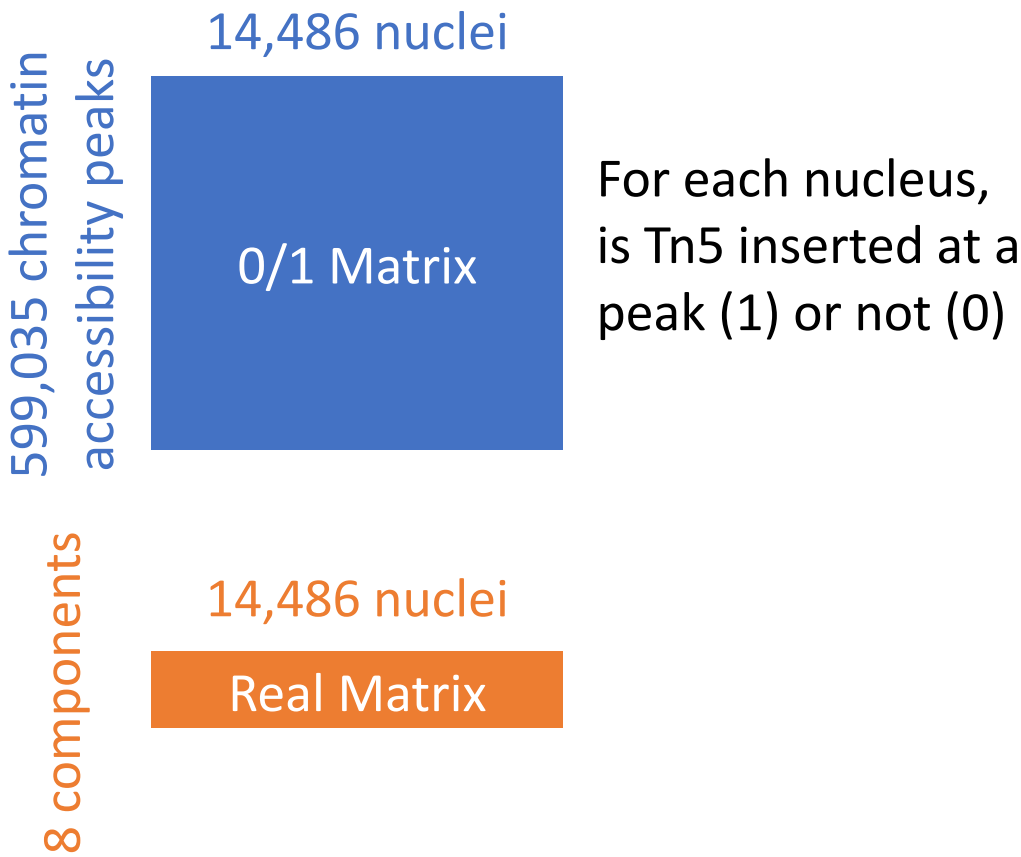
1. High-dimensional input data

Dimension reduction

2. Low-dimensional space representing biology

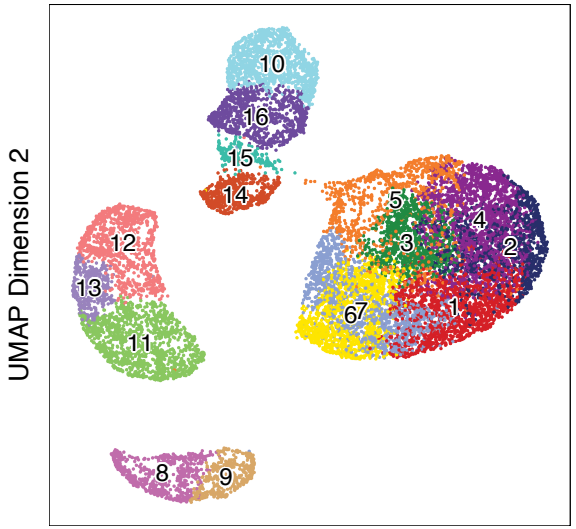
Clustering

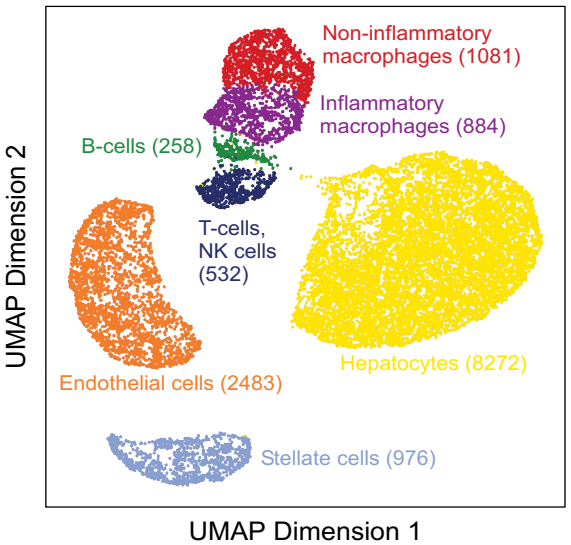
3. Classify nuclei



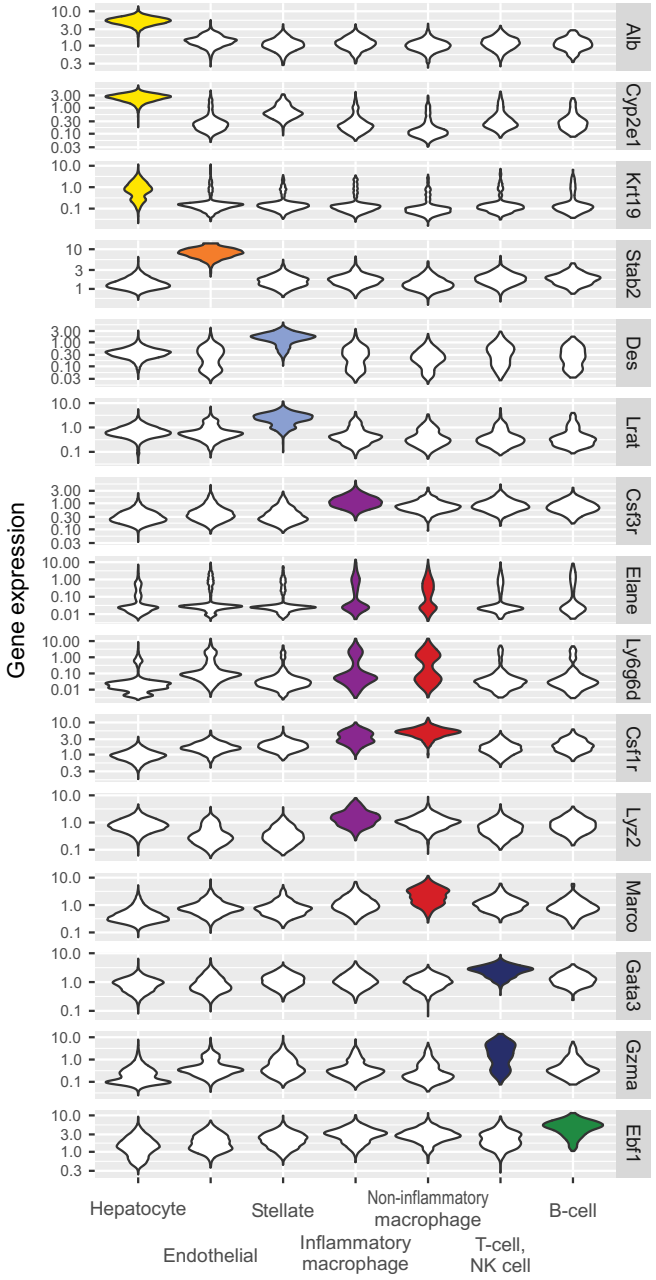
Observed cell types

- 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



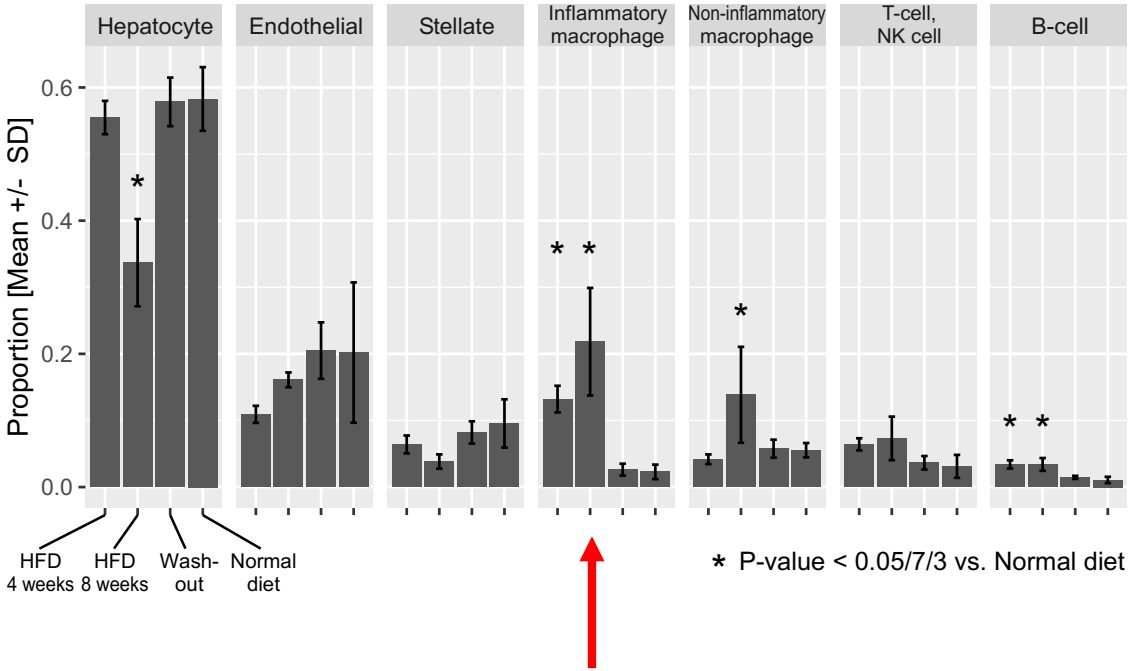


3000 nuclei per sample



Cell type composition

- Methods: Inferred cell type composition in bulk tissue samples (n=4 per condition)
  - Measured bulk ATAC-seq profile
  - Approximated as a mixture of profiles for pure cell types (obtained by snATAC-seq)
- 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



Agree with steatohepatitis after 8 weeks HFD

Differential gene expression by diet, in each cell type

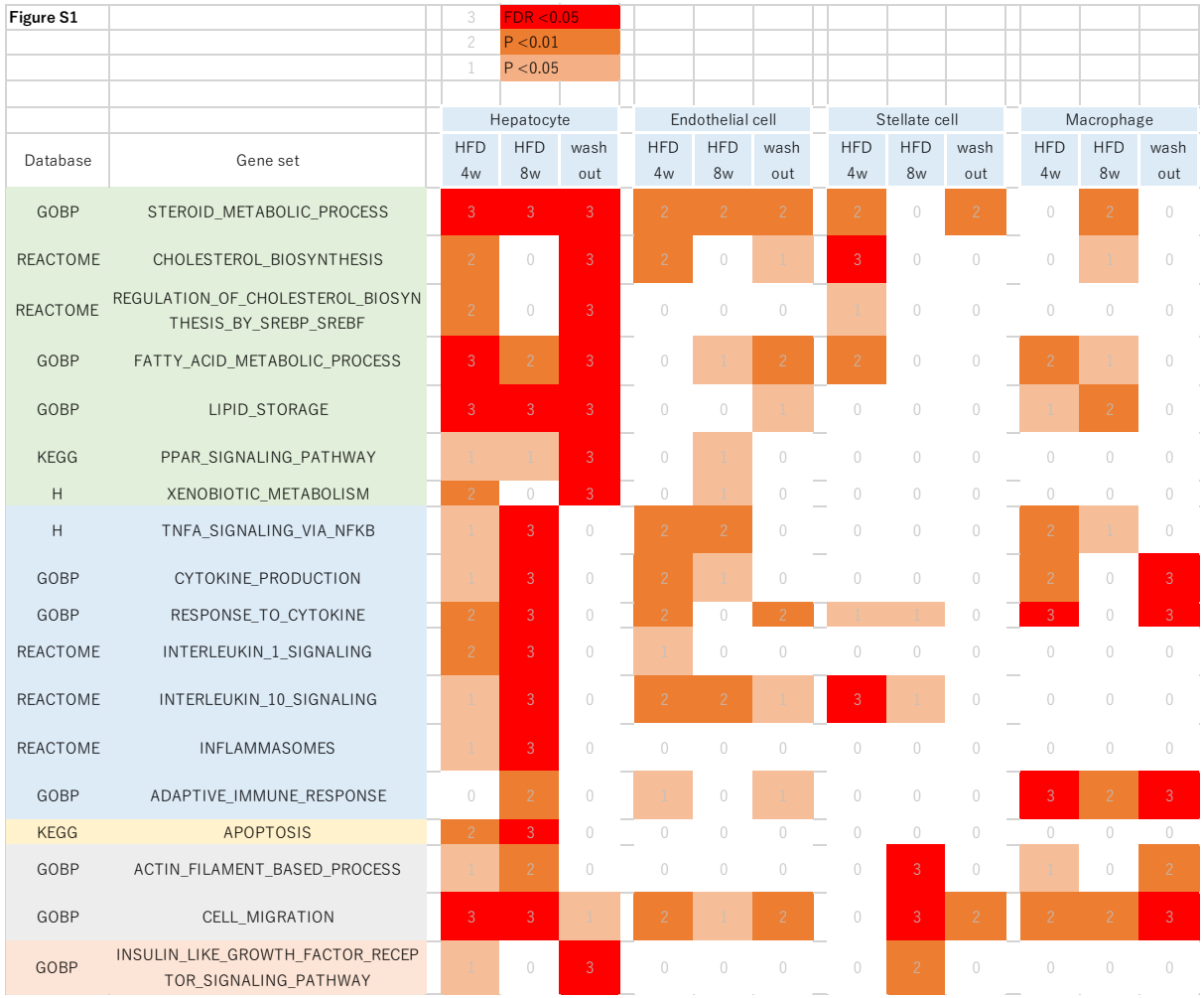


Figure S1		3	FDR <0.05												
		2	P <0.01												
		1	P <0.05												
			Hepatocyte			Endothelial cell			Stellate cell			Macrophage			
Database	Gene set		HFD 4w	HFD 8w	wash out	HFD 4w	HFD 8w	wash out	HFD 4w	HFD 8w	wash out	HFD 4w	HFD 8w	wash out	
GOBP	STEROID_METABOLIC_PROCESS		3	3	3	2	2	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	

Compared to normal diet

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

Agree with steatohepatitis after 8 weeks HFD

Summary (1)

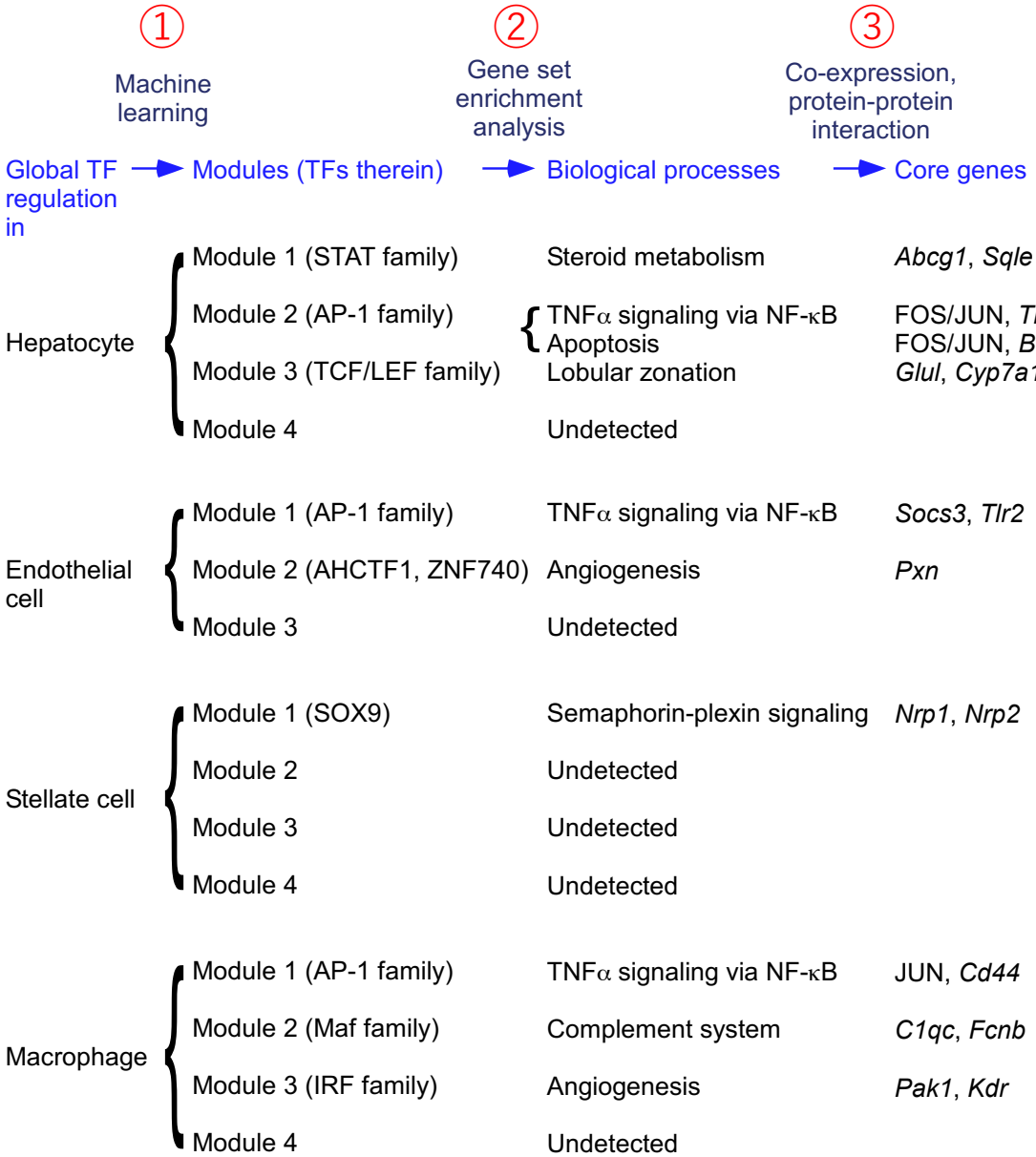
- By utilizing single-nucleus and bulk ATAC-seq, we could observe cell-type specific changes in the development of MASLD.
  - Changes in cell type composition
  - Changes in cell-type-specific gene expression
  - Both agreed with the pathological progression

Part 1. Development of MASLD

- Quantity: cell type composition
- Quality:
  - gene expression
  - gene regulation

Part 2. Recovery from MASLD

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 → Modules (TFs therein)

For each gene, compute its regulator TFs/diets

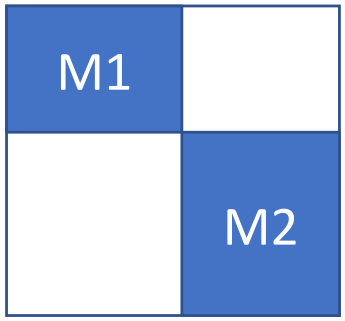
541 TFs + 3 diets

8,827 genes

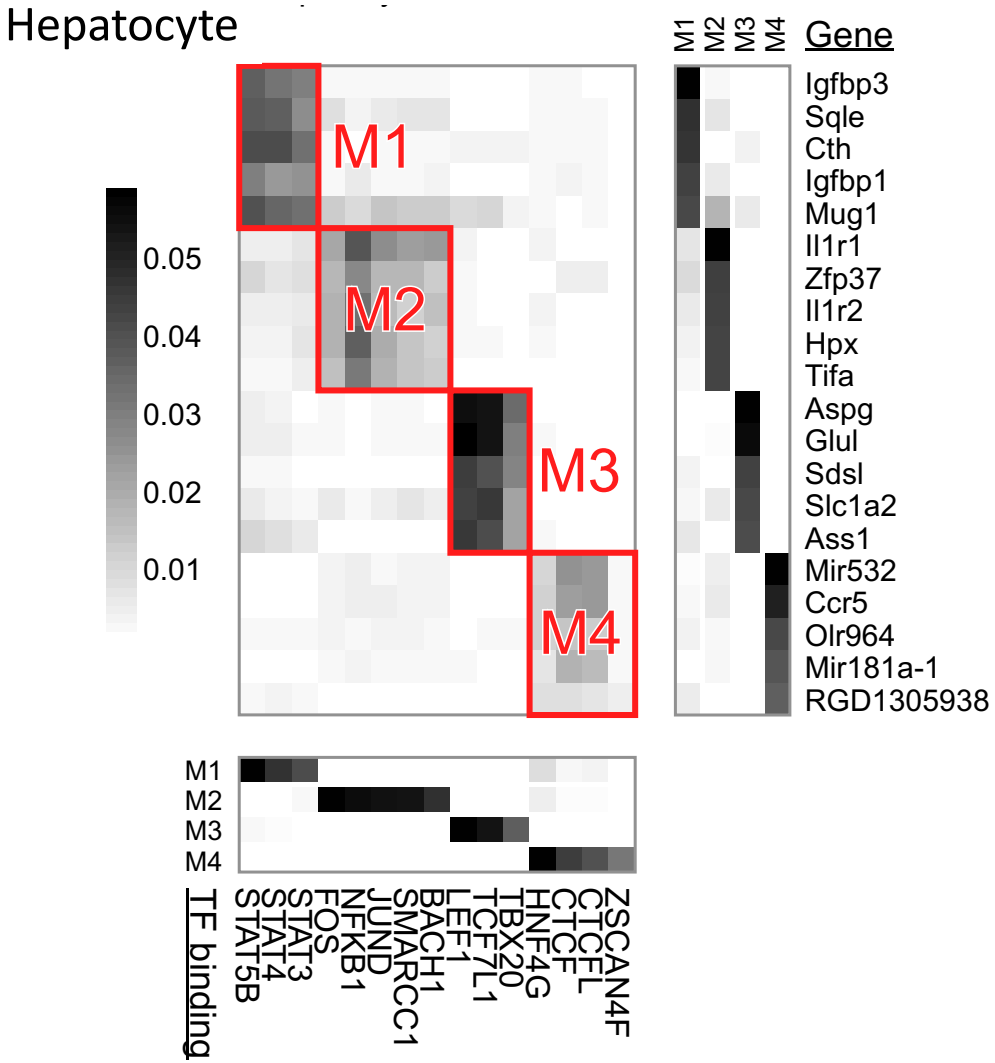
Regulator-Regulatee Matrix

Extract modules

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



Major modules of TF regulation



Four modules in hepatocytes

Only representative TFs and genes are shown.

① Machine learning

② Gene set enrichment analysis

③ Co-expression, protein-protein interaction

Modules (TFs therein) → Biological processes

- Essentially, a module is a list of TFs and genes
- For the list of genes, find characteristic biological processes
  - Search in database for processes whose Gene Set overlaps significantly

Biological processes characterizing TF modules

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 agreed with literature for 5 out of 7
- Good indication!

① Machine learning

② Gene set enrichment analysis

③ Co-expression, protein-protein interaction

Biological processes → Core genes

- A biological process is defined as a gene set (GS)
- Core genes of a GS
  - Central in co-expression
  - Central in protein-protein interaction
    - STRING database

Core genes found in this study

Cell type	Biological process	#Core genes	Known causal or biomarker genes for MASLD	
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 MASLD genes
- Suggests the biological validity of our data-driven approach

Summary (2)

- Using novel statistical methods,
- 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 literature.
- We searched core TFs/genes in biological processes, many of which overlapped with known MASLD genes.

Part 1. Development of MASLD

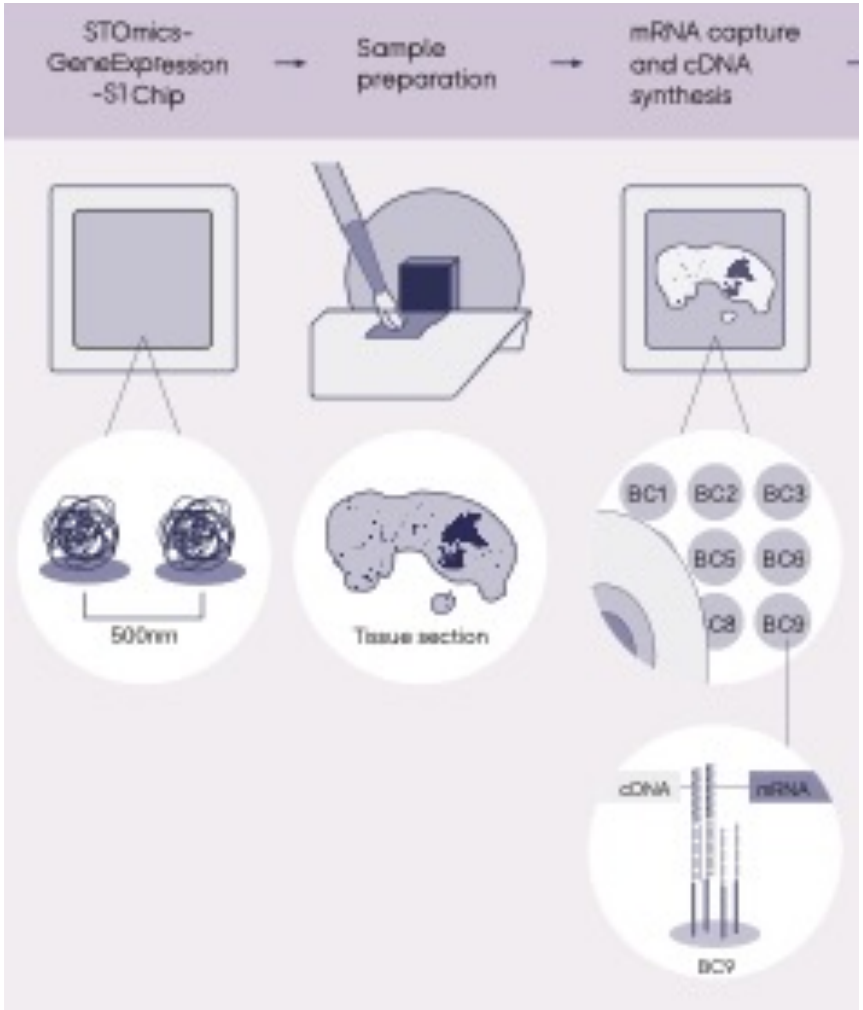
- Quantity: cell type composition
- Quality: gene expression and regulation

Part 2. Recovery from MASLD

- Distinguish recovering cells
- Single-cell + Spatial transcriptomics

Spatial transcriptomics

- Put a frozen tissue section on a 1cm x 1cm chip
  - RNA-seq on 0.5 μm spaced spots
  - Binned 50 x 50 spots (25 μm resolution)
- We used Stereo-seq
- Cons: cells not demarcated
  - Borders can be inferred
- Complementary to single-cell assays, but not replacement



BGI Research

Background & Aim

- Epidemiologically, weight loss by diet or exercise is proven to ameliorate MASLD/MASH.
- Cellular and molecular mechanism is unclear.
- In animals, MASH development has been studied, but not recovery.
  - Diet-based recovery model not reported

- Understand gene expression in the recovery process from MASH
- Distinguish the “recovering cells”

➔ Single-cell + spatial transcriptomics

Summary (3)

- By integrating single-cell and spatial transcriptomics assays, we identified which cell types exist in the recovering and non-recovering regions of liver tissue.

- Ongoing
  - Spatial change in gene expression within each cell type
  - Cell trajectory during recovery or non-recovery (e.g. fibrosis)

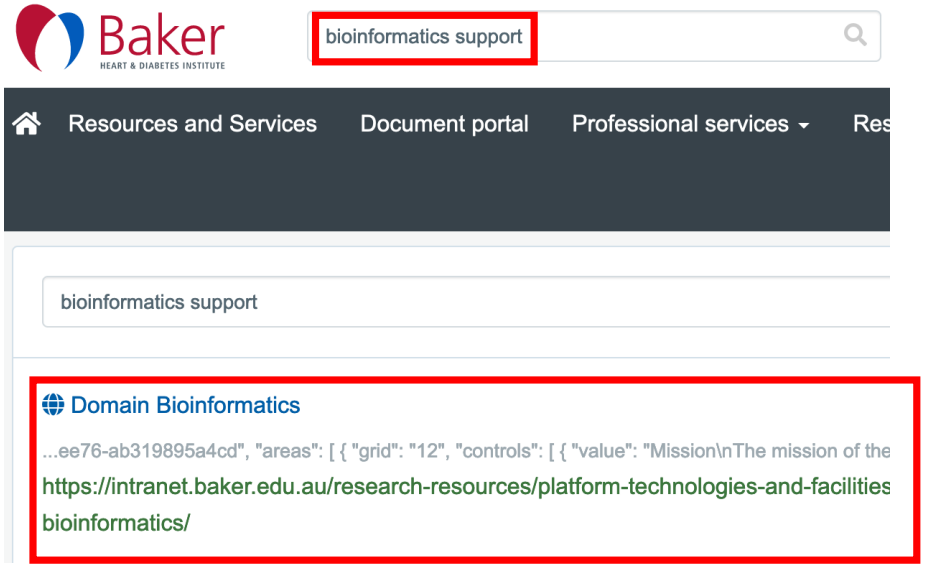
https://www.fumihiko.takeuchi.name

Supporting bioinformatics for all at Baker

Popular collaborations

- UK Biobank (n=500K) analysis
- Omics
  - Genomics
  - Transcriptomics
  - Single-cell
  - Spatial transcriptomics
  - etc.

- Domain Bioinformatics support page
  - Search in Baker intranet



- Bioinformatics AT baker.edu.au