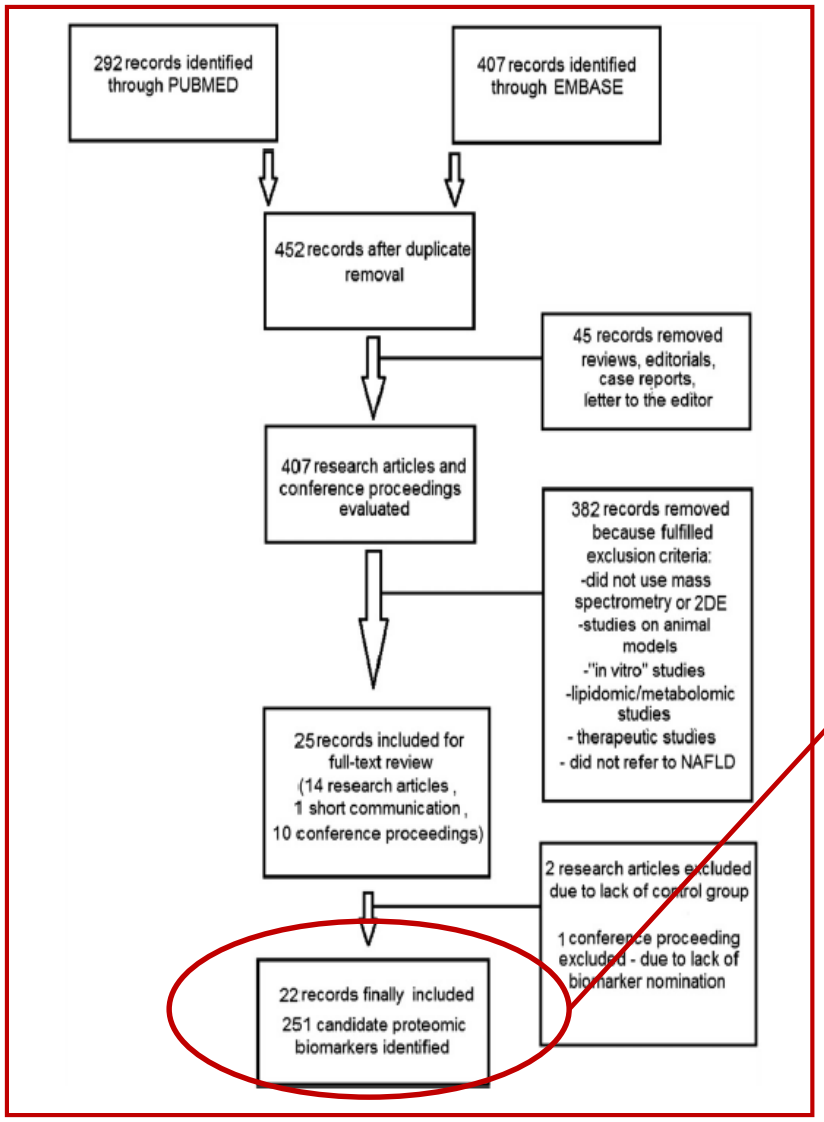
Proteomics based biomarker discovery in NAFLD

Naga Chalasani, MD
David W. Crabb Professor& Director
Division of Gastroenterology and Hepatology
Indiana University School of Medicine

Biomarkers in NASH – Unmet and urgent needs

- Diagnostic biomarkers: Which NAFLD patients have NASH, NASH with F2-F3, or Cirrhosis?
- Biomarkers to replace liver biopsy and HVPG in registration clinical trials
- Prognostic biomarkers:
  - To identify NAFLD patients are at risk for cirrhosis?
  - To identify NAFLD cirrhotic patients at risk for decompensation

Proteomics & NAFLD/NASH: Systematic Review



- 8 studies based on liver biopsy samples
- 1 stool based study
- 13 serum based – several comparing NAFLD vs. normal and several with small sample sizes

Ladaru A, et al. Biomarkers 2016

Serum proteomics & biomarker discovery across the spectrum of NAFLD

- 69 subjects with varying stages of NAFLD and 16 controls
- Label-free mass spectrometry
- Chicken lysozyme was spiked into each sample at a constant amount as the internal reference
- Identified 1700 protein with a peptide identification confidence >75%
- 605 proteins were different between any two groups with FDR < 5%. Fold change between groups was calculated

Bell LN, et al. Hepatology 2010

Table 1. Clinical Characteristics and Liver Biochemistries of Study Participants

	Controls n=16	Simple Steatosis n=24	NASH n=23	NASH F3/F4 n=22
Age (years)	43 ± 14	50 ± 12	44 ± 12	52 ± 11
BMI (kg/m^2)	32.2 ± 3.8	32.6 ± 8.9	33.8 ± 6.2	35.4 ± 8.8
Male (%)	44	71	39	32
Caucasian (%)	100	83	96	95
Comorbidities				
Hypertension (%)	12.5	50	50	74
Diabetes mellitus (%)	0	31	40	58
Dyslipidemia (%)	12.5	47	39	50
Liver biochemistries				
AST (U/L)	22 ± 8	53 ± 37	66 ± 41	70 ± 34
ALT (U/L)	25 ± 8	63 ± 29	89 ± 55	70 ± 40
ALP (U/L)	74 ± 17	98 ± 38	99 ± 56	93 ± 49
Total bilirubin (mg/dL)	0.4 ± 0.2	0.9 ± 0.6	0.6 ± 0.2	0.8 ± 0.3

Values are expressed are mean ± SD.

BMI, body mass index; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase.

Five NAFLD patients had methotrexate use, three patients in the simple steatosis group and two patients in the NASH group.

Table 2. Summary of All Identified Proteins

Protein Priority	Peptide ID Confidence	Multiple Sequences Quantified	Number of Proteins	Number of Significant Changes*	Maximum Absolute Fold Change	Median % Coefficient of Variation
1	High	Yes	183	72	2.55	16.85
2	High	No	426	148	14.26	26.33
3	Moderate	Yes	73	30	2.28	20.42
4	Moderate	No	1056	355	14.22	26.39
		Overall	1738	605	14.26	24.73

*False discovery rate (FDR) <5% ($q < 0.05$).

Table 3. Pairwise Summary of Significant Changes Among All Four Patient Groups

		Number of Significant Changes Between Groups ($q<0.05$)						
Protein Priority	Number of Proteins	Control Versus Simple Steatosis	Control Versus NASH	Control Versus NASH F3/F4	Control Versus All NAFLD/NASH	Simple Steatosis Versus NASH	Simple Steatosis Versus NASH F3/F4	NASH Versus NASH F3/F4
1	183	30	36	47	20	0	21	9
2	426	90	107	97	59	0	11	1
3	73	16	21	20	12	0	2	0
4	1056	187	248	253	138	0	21	5
Overall	1738	323	412	417	229	0	55	15

56 priority 1 proteins had a significant change between groups >14% (q<0.05). 14% was maximum observed change for the internal reference

Table 4. Summary of Biological Processes in Which Differentially Expressed Priority 1 Proteins with a Significant Change >14% ($q < 0.05$) Are Involved

Biological Process	Number of Proteins	Protein List
Immune system regulation and inflammation	15	α-1-acid glycoprotein 2, complement component C7, α-1-acid glycoprotein 1, serum amyloid P component, α-2-macroglobulin, CD5 antigen-like, C-reactive protein (isoform 1), complement C4A, complement component 4B (preprotein), complement component 4A, N-acetylmuramoyl-L-alanine amidase (isoform 1), complement comp 1q subcomp (β chn precursor), GUGU (β form), fetuin-B, complement C1s (subcomponent)
Coagulation cascade	14	Platelet basic protein, fibrinogen β chain, fibrinogen γ chain (isoform γB), SERPINC1, antithrombin III variant, platelet factor 4, platelet factor 4 variant, prothrombin (fragment), α-2-macroglobulin, histidine-rich glycoprotein, von Willebrand factor, SERPINF2, fibrinogen α chain (isoform 1), plasminogen
Structural and extracellular matrix proteins	9	Transgelin 2, actin (cytoplasmic), serum amyloid P component, lumican, proteoglycan 4 (isoform A), transthyretin, extracellular matrix protein 1, α-2-glycoprotein 1 (zinc), gelsolin (isoform 1)
Blood carrier proteins	7	Retinol binding protein 4, insulin-like growth factor acid labile subunit, transthyretin, insulin-like growth factor binding protein 2, insulin-like growth factor binding protein 3, afamin, fetuin-B
Cholesterol and triglyceride balance (lipoprotein components)	7	Apolipoprotein C1, apolipoprotein A2, apolipoprotein A4, apolipoprotein A4 precursor, apolipoprotein B100, apolipoprotein C3, apolipoprotein L1 (isoform 2)
Cell growth/survival/proliferation	6	Insulin-like growth factor acid labile subunit, insulin-like growth factor binding protein 2, insulin-like growth factor binding protein 3, sterile α motif domain-containing protein 9 (isoform 1), α-2-glycoprotein 1 (zinc), gelsolin (isoform 1)
Antiinflammatory and anti-oxidant	5	Paraoxonase 1, retinol binding protein 4, apolipoprotein A2, afamin, apolipoprotein L1 (isoform 2)
Gene/protein processing and expression	3	Poly (A) RNA polymerase mitochondrial (isoform 1), replication initiation-like protein (isoform 1), sulfhydryl oxidase 1 (isoform 1)
Unknown	2	13 kDa protein, 11 kDa protein
Blood pressure regulation (renin-angiotensin system)	1	Angiotensinogen
Insulin action	1	Retinol binding protein 4
Pro-oxidant	1	Sulfhydryl oxidase 1 (isoform 1)

Biomarker discovery

- 27 proteins with fold change >30% between groups were further chosen for biomarker discovery
- To assess the diagnostic utility, combinations of proteins were used to identify 3 different classification groups – (a) distinguish all four groups; (b) distinguish NAFL from NASHF3/F4; and (c) distinguish all NAFLD from controls

Group 1: 27 proteins for classifying all 4 groups

Annotation	Number Classified Correctly* (%)	Area Under the ROC Curve			
		Control	Simple Steatosis	NASH	NASH F3/F4
Platelet basic protein	41 (48%)	0.98	0.71	0.72	0.75
Fibrinogen β chain	39 (46%)	0.96	0.67	0.71	0.69
Fibrinogen γ chain (isoform γB)	38 (45%)	0.94	0.65	0.64	0.68
Poly (A) RNA polymerase, mitochondrial (isoform 1)	37 (44%)	0.89	0.54	0.63	0.65
SERPINC1	31 (36%)	0.87	0.59	0.65	0.6
Antithrombin III variant	30 (35%)	0.87	0.59	0.63	0.59
Paraoxonase 1	30 (35%)	0.81	0.56	0.53	0.69
Platelet factor 4	40 (47%)	0.88	0.72	0.66	0.69
Platelet factor 4 variant	37 (44%)	0.87	0.72	0.65	0.72
Prothrombin (fragment)	40 (47%)	0.96	0.63	0.73	0.68
α-1-acid glycoprotein 2	32 (38%)	0.62	0.63	0.7	0.74
Retinol binding protein 4	29 (34%)	0.58	0.6	0.6	0.78
Transgelin 2	34 (40%)	0.72	0.62	0.66	0.66
Apolipoprotein C1	35 (41%)	0.49	0.59	0.73	0.74
Replication initiation-like protein (isoform 1)	30 (35%)	0.71	0.6	0.54	0.46
Complement component C7	30 (35%)	0.74	0.68	0.57	0.8
α-1-acid glycoprotein 1	34 (40%)	0.65	0.65	0.66	0.7
Actin (cytoplasmic)	27 (32%)	0.65	0.59	0.59	0.6
Serum amyloid P component	34 (40%)	0.47	0.61	0.66	0.8
α-2-macroglobulin	32 (38%)	0.76	0.64	0.63	0.68
13 kDa protein	29 (34%)	0.72	0.61	0.58	0.71
Lumican	31 (36%)	0.87	0.58	0.61	0.74
Insulin-like growth factor acid labile subunit	29 (34%)	0.57	0.52	0.62	0.69
Histidine-rich glycoprotein	28 (33%)	0.79	0.51	0.64	0.56
Proteoglycan 4 (isoform A)	31 (36%)	0.45	0.48	0.67	0.7
CD5 antigen-like	28 (33%)	0.62	0.55	0.48	0.66
von Willebrand factor	34 (40%)	0.86	0.58	0.63	0.62

*Of 85 total subjects.

Group 2: 27 proteins for differentiating SS/NASH from NASH with F3/F4classifying all 4 groups

Annotation	Number Classified Correctly* (%)	Area Under the ROC Curve: NAFLD vs NASH F3/F4
Platelet basic protein	52 (75%)	0.75
Fibrinogen β chain	43 (62%)	0.6
Fibrinogen γ chain (isoform γB)	37 (54%)	0.59
Poly (A) RNA polymerase, mitochondrial (isoform 1)	38 (55%)	0.56
SERPINC1	35 (51%)	0.5
Antithrombin III variant	31 (45%)	0.49
Paraoxonase 1	44 (64%)	0.63
Platelet factor 4	48 (70%)	0.72
Platelet factor 4 variant	46 (67%)	0.73
Prothrombin (fragment)	42 (61%)	0.68
α-1-acid glycoprotein 2	48 (70%)	0.78
Retinol binding protein 4	48 (70%)	0.78
Transgelin 2	41 (59%)	0.62
Apolipoprotein C1	46 (67%)	0.75
Replication initiation-like protein (isoform 1)	36 (52%)	0.52
Complement component C7	50 (72%)	0.76
α-1-acid glycoprotein 1	49 (71%)	0.75
Actin (cytoplasmic)	37 (54%)	0.57
Serum amyloid P component	50 (72%)	0.82
α-2-macroglobulin	46 (67%)	0.63
13 kDa protein	45 (65%)	0.68
Lumican	42 (61%)	0.67
Insulin-like growth factor acid labile subunit	43 (62%)	0.69
Histidine-rich glycoprotein	32 (46%)	0.48
Proteoglycan 4 (isoform A)	44 (64%)	0.73
CD5 antigen-like	44 (64%)	0.64
von Willebrand factor	40 (58%)	0.53

*Of 69 total subjects.

Biomarker discovery

- Of 27 proteins with >30% fold change between groups, 10 proteins with a high % confidence that are able to differentiate between groups

Table 5. Protein Identification Confidence of Potential Biomarker Candidates

Protein ID	Annotation	Minimum *q*-value	Number of Amino Acid Sequences ID'd	Protein ID Confidence (%)	Sequence Coverage* (%)
IPI00298497.3	Fibrinogen β chain	0.000000	17	100	52
IPI00218732.3	Paraoxonase 1	0.000000	11	100	55
IPI00019568.1	Prothrombin (fragment)	0.000000	23	100	53
IPI00022420.3	Retinol binding protein 4	0.000001	9	99.99	43
IPI00550363.3	Transgelin 2	0.000000	2	99.99	18
IPI00296608.6	Complement component C7	0.000007	22	100	42
IPI00022391.1	Serum amyloid P component	0.000000	12	100	75
IPI00020986.2	Lumican	0.000028	10	99.99	41
IPI00020996.3	Insulin-like growth factor acid labile subunit	0.000000	6	100	17
IPI00025204.1	CD5 antigen-like	0.006677	3	99.99	14

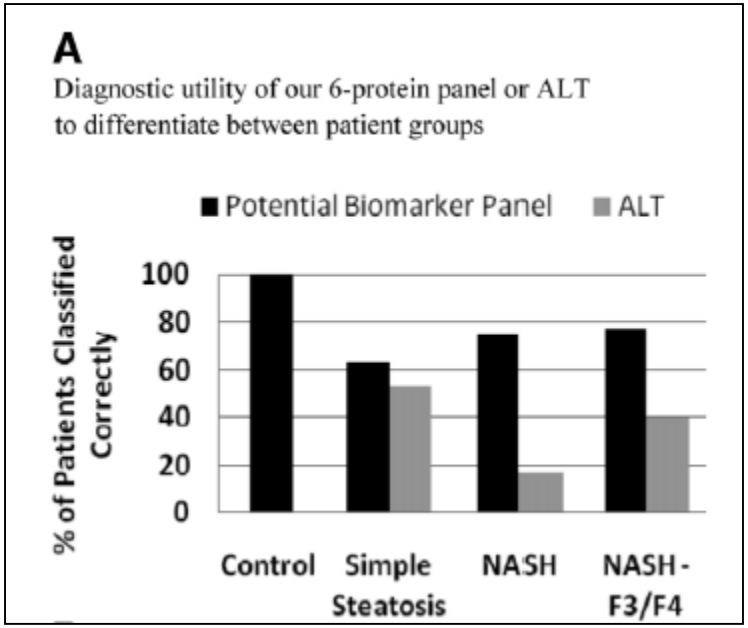
*Percent of the complete protein amino acid sequence where matching peptides for protein identification were found.

Biomarker panels

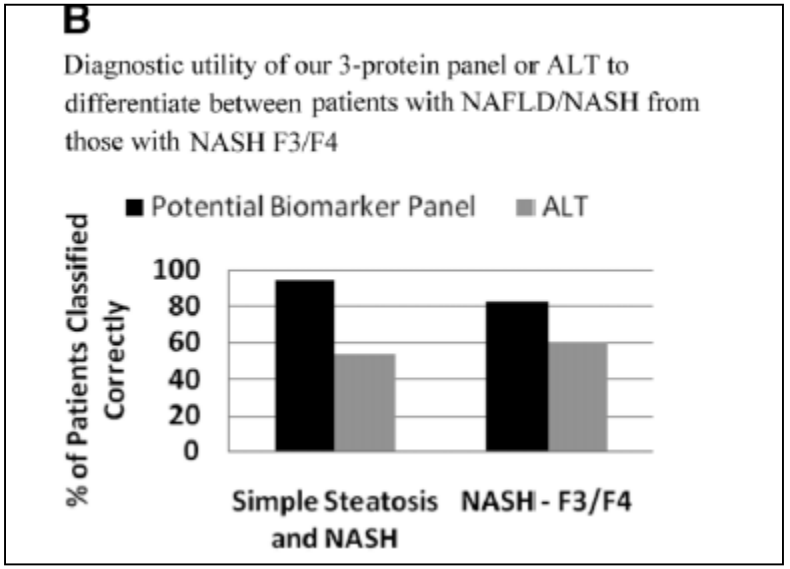
- **6-protein panel** *(fibrinogenβ chain, RBP4, serum amyloid P component, lumican, transgelin 2, and CD-24 antigen like)* differentiates all four groups with success rate of 76%
- **3 protein panel** *(complement component C7, insulin-like growth factor acid labile subunit, and transgelin2)* has 90% success in differentiating NASH with F3/F4 from NAFLD without F3/F4
- **2 protein panel** *(prothrombin fragment and paroxonase 1)* 100% success rate in differentiating NAFLD and controls

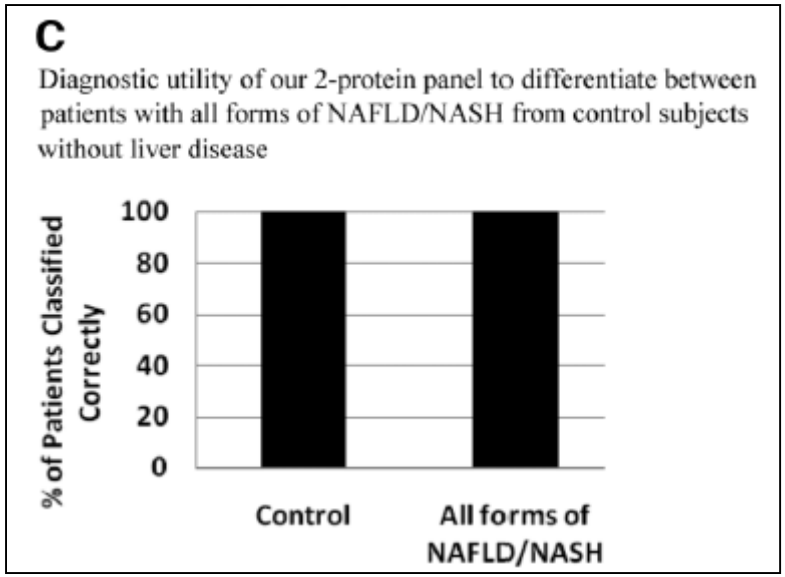
AUROCs for biomarker panels

	Control	Simple steatosis	NASH	NASH F3/F4
6 protein panel	1.0	0.83	0.86	0.91
ALT		0.68	0.59	0.69
3 Protein panel		0.91 to distinguish SS and NASH from NASH with F3/F4		
ALT		0.53 to distinguish SS and NASH from NASH with F3/F4		
2 protein panel	1.0	1.0 to distinguish NAFLD from control		



Panels perform better than ALT in distinguishing various NAFLD subgroups





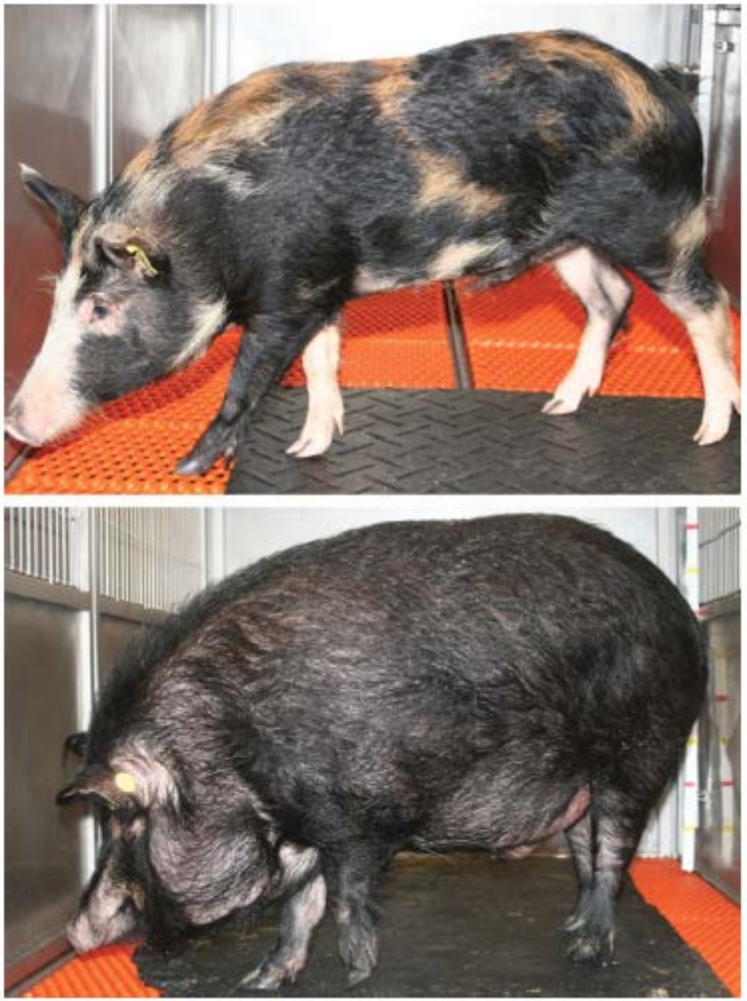
Bell LN, et al. Hepatology 2010

Serum proteomics in the Ossabaw NASH model

Table 1. *Characteristics of Ossabaw miniature pigs at time of euthanasia*

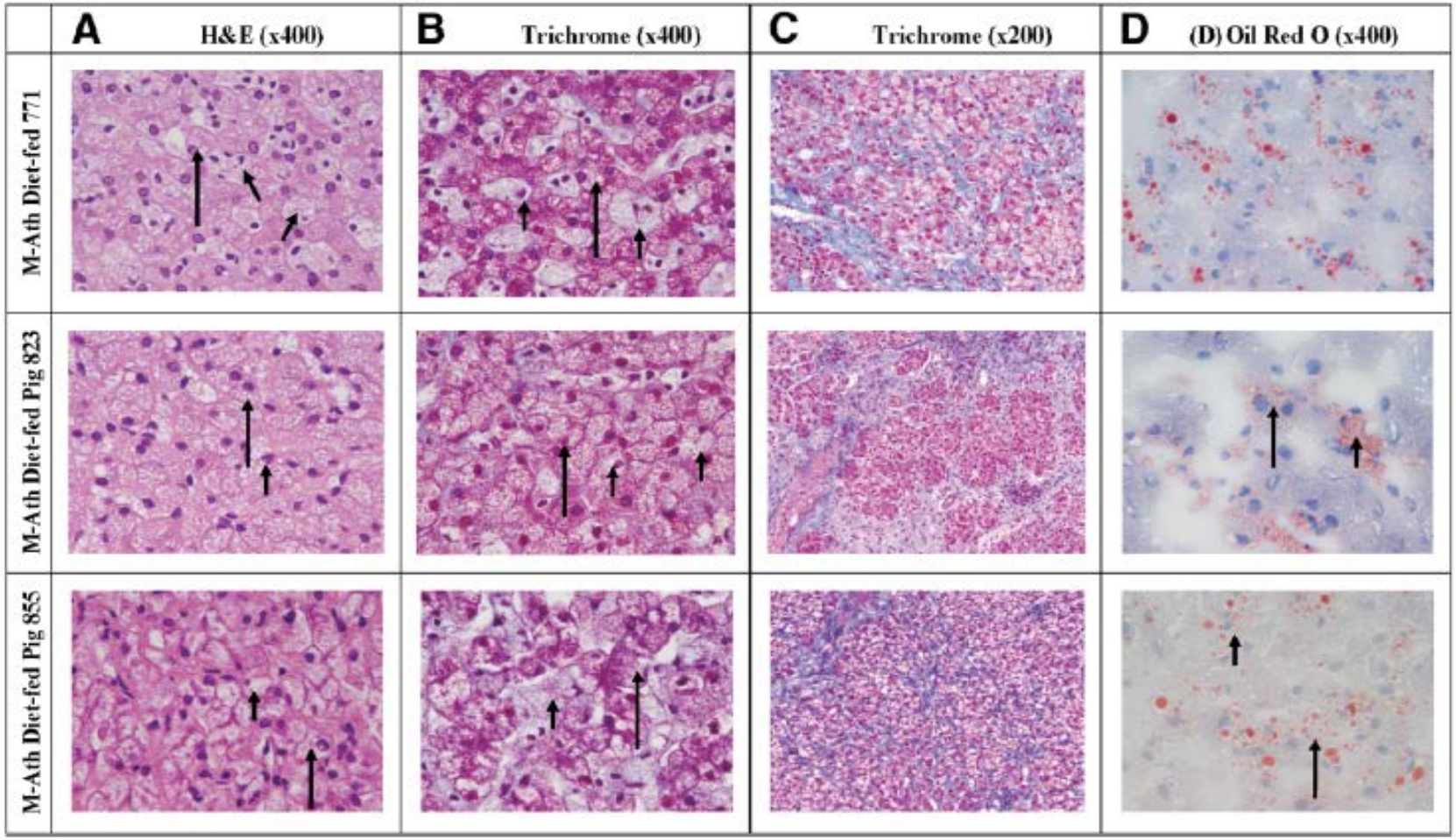
	Control Chow Group ($n = 7$)	Atherogenic Diet Group ($n = 5$)	M-Ath Diet Group ($n = 6$)
Sex, male/female	2/5	0/5	0/6
Weight at euthanasia, kg	53 ± 2	94 ± 4*	86 ± 14*
Mean weight gain, kg	14 ± 1	54 ± 4*	38 ± 13
Glycemic measures			
Fasting glucose, mg/dl	79 ± 3	87 ± 2	88 ± 6
Fasting insulin, μU/ml	11 ± 2	15 ± 1	18 ± 3
HOMA	2 ± 0.3	3 ± 0.3	4 ± 1
Lipid measures			
Total cholesterol, mg/dl	81 ± 4	328 ± 32*	629 ± 80*†
LDL cholesterol, mg/dl	36 ± 4	242 ± 27*	520 ± 68*†
HDL cholesterol, mg/dl	41 ± 4	79 ± 6*	83 ± 5*
LDL/HDL Ratio	1 ± 0.2	3 ± 0.3*	6 ± 1*†
Triglycerides, mg/dl	24 ± 3	37 ± 2	130 ± 17*†
Liver biochemistries			
AST, U/l	32 ± 3	27 ± 1	100 ± 21*†
ALT, U/l	46 ± 7	33 ± 1	41 ± 12
ALP, U/l	73 ± 11	110 ± 10	273 ± 110*
Total bilirubin, mg/dl	0.2 ± 0.02	0.2 ± 0.02	0.3 ± 0.04

Values expressed are means ± SE. *$P < 0.05$ vs. control chow group. †$P < 0.05$ vs. atherogenic diet group. M-Ath, modified atherogenic; HOMA, homeostatic model assessment method; LDL, low-density lipoprotein; HDL, high-density lipoprotein; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase.

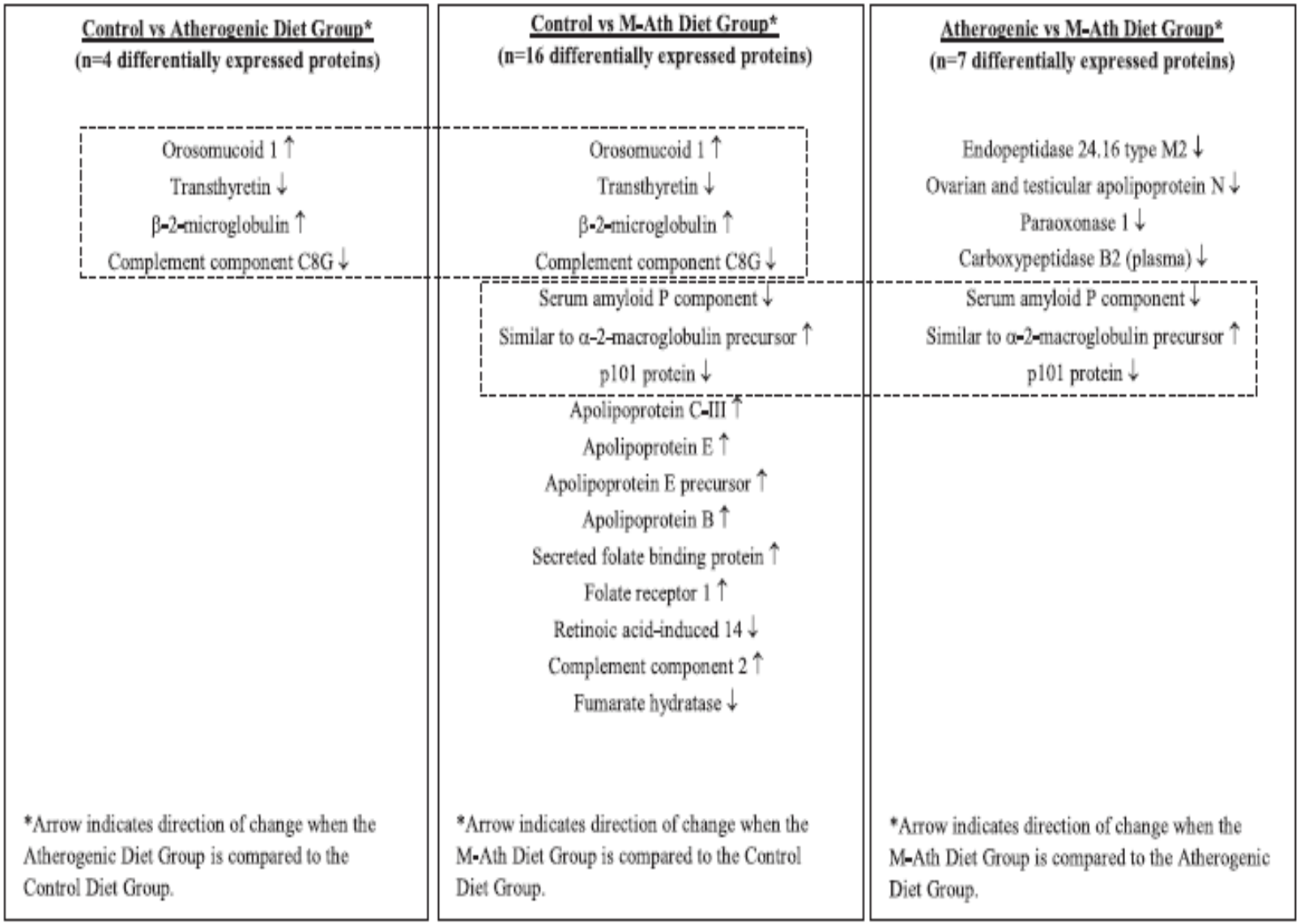


Bell LN, et al. AJP-GI 2010

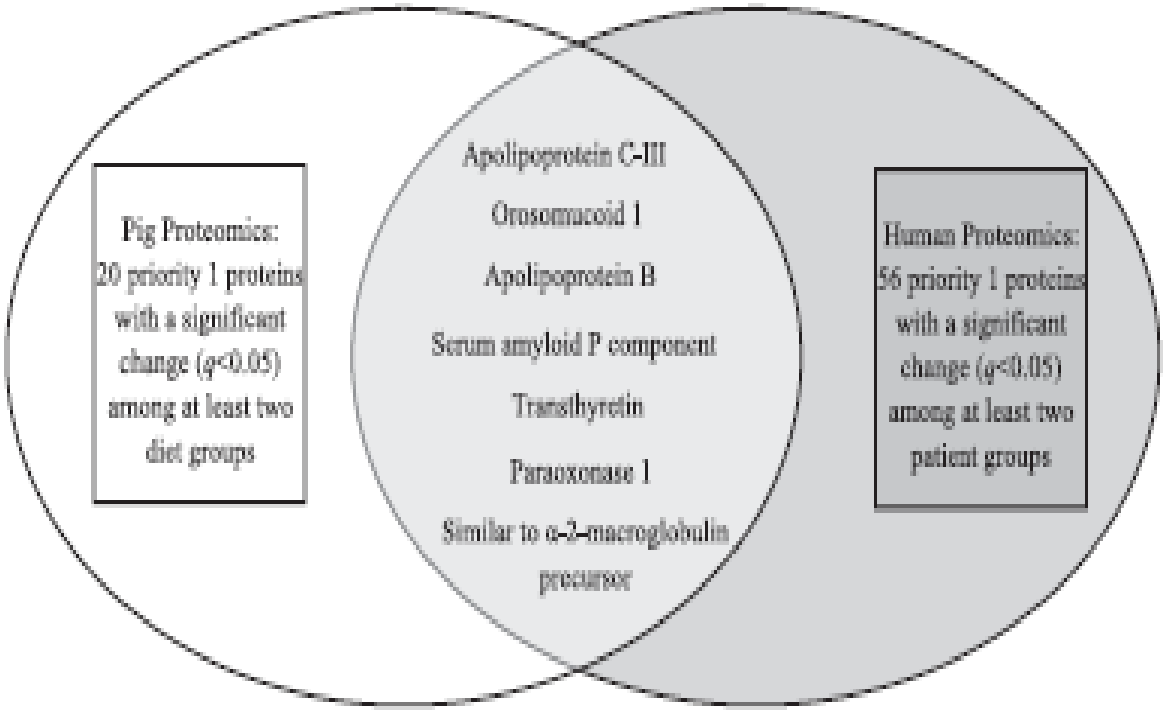
Metabolic liver Injury in the Ossabaw



20 Priority 1 proteins are differentially expressed between groups



Shared proteins between human and pig proteomic studies



Bell LN, et al. AJP-GI 2010

Summary

- Proof-of-concept studies done to date suggest that proteomics-based biomarker discovery and validation is a promising strategy
- Multidisciplinary team of hepatologists with phenotype expertise and access to biosamples, biochemists with technical expertise and informaticians
- Could be combined with other “Omics” and/or clinical and laboratory variables