Work package 4b CatApp Toxicogenomics data where does it help?

Cat-App final event, Brussels, 6 September 2018

Shu-Dong Zhang, University of Ulster, Northern Ireland



Reproduction permitted with due © Concawe acknowledgement

## TempO-seq Technology - High-Throughput Targeted Sequencing



#### Assay advantages:

- Works in 384-wells
- No cDNA library prep
- 1000's genes/sample
- High Specificity (probe seq + ligation)

#### Assay considerations:

- Sequencing depth (per gene per sample)
- Gene selection (targeted set vs whole genome)
- New technology (no large database for comparisons)



#### Cat-App Transcriptomics Data - 6 Human Cell Types





11,000+ samples from 6 cell types

- 4-point concentration response data
- Differential gene expression for ~3,000 transcripts (targeted analysis)
- Over 35,000,000 data points
- Novel data processing pipeline
- Concentration-response modeling pipeline
- Transcriptomics data can be combined with other data streams

#### Transcriptomic Data Analysis - Pipeline



House et al. Front Genet 8:168, 2017

### Transcriptomic Data Analysis - Quality Control

- Raw reads are demultiplexed and mapped
- Minimum read
  count is set per gene -
- Examination of controls (DMSO, Method blank, media)
- Examination of sequencing library quality



#### Transcriptomic Data Analysis - Effect of Petroleum Substances



#### Transcriptomic Data Analysis - Effect of Petroleum Substances





Number of Differentially Expressed Genes

#### Cell-specific Gene Expression Signature Across all Substances



#### iCell Hepatocytes



Number of petroleum substances that "perturbed" each transcript [top 50 shown]

Direction

Down

Up

75

25

50

#### Hepatocyte Gene Expression: Group-Specific "Signatures"?

UATO	AE	HFOC	0_06	RGO	индо	BIT	KER	во	NAPTI	HWAX	P.LAT									DIT					
			1					1	1					TIPAR		AE	100_00	JU SKGO	VHGO	BII	KER	BO	NAPTHA	WAX	P.LAI
							1.1	1						NOO	1 -							i			
							£							IGFBP	4-							i			
					F.			ł.				Ν		СҮРЗА	5 -							î 👘	1		
								1						UGT1A	3 -										
					E									SOS	1 -										
					E			1	1					CAV	1 -						_	_	_		
		E.			F			1						CYP1A	1										
		Ē.			1		E .							HERPUD	<u></u>							5 - E			
					÷ .									CDH	41							2 - C			
			E I	-	E									I MEM9								2			
		2	E											SIC13A	[]										
								1						TSK	í]							1			
		E.	-		1									ABCC	5-							i			
		E.		_										H BM	-						1.1	i			
		1			E			-	-					ថ្ថិ ORM	2 -							Ĩ.			
					1		1 - I								1-							Ĩ			
			Ē.		÷		-		_			<u>۱</u>		8FE2L	2 -								1		
-					8				1			<u>۱</u>		WDR1	2		 _		_	_					
<b>_</b>			F I					1				\		EGFBP	1-							L			
-		E.						i.						0 MYL	<u></u>										
<b>—</b>		L.					1						1		<u>}</u>							1			
			÷ 1										1		51					<b>.</b>					
	1 - I		E		=				1				1		3]										
					÷		=	ł.					1	0 TAS							1.1	i			
													1	HSD17B1	1 -						- C	î			
					-								1	GNM	÷-							i –			
					-			1					1	ALAS	1 -							i			
	2	1			-				_				1	AKR1C	2 -							i			
					1								1	AKR1C	3 -							1			
	Ξ							ł.					· · ·	PSMB	3 -		_		_						
			-										· · ·	SOCS	2 -										
	-						2						· \	TAGL	N -							Į			
	-		i					1	-				1	NFKBI	21										
			÷					1	-				1	HSD11B	<u> </u>										
													1		5										
													<u>۱</u>	APBB ETC	<sup>2</sup> ]				-			1			
														CAD3V	4 -						1.1	1 C	1.1		
																					. <b>1</b>				
														-										100	Same Stre

#### Cardiomyocyte Gene Expression: Group-Specific "Signatures"?



## Hepatocyte Gene Expression: Group-Specific "Pathways"?

			VHGO	CGO and OGO	Aromatic Exracts	HFO & FO	UATO	
		Steroid hormone biosynthesis	FDR=2.5e-6	FDR=8.6e-8	FDR=1.0e-4	FDR=3.9e-8	FDR=8.2e-7	7
CVD/FO avidation	5	Metabolism of xenobiotics by cytochrome P450	FDR=7.9e-6	FDR=1.6e-7	FDR=3.9e-4	FDR=4.9e-6	FDR=4.7e=6	
	$\leq$	Retinol metabolism	FDR=2.8e-4	FDR=2.6e-6	R=5.6e-3	FDR=5.8e-4	FDR=4.0e-4	
		Biological oxidations	R=6.1e-3	FDR=8.1e-5	DR=3.0e-3	FDR=5.8e-4	FDR=4.0e-4	
		PERk regulated gene expression Drug metabolism – other enzymes	DR=1.5e-3		=1.3e=2	DR=1.7e-3	FDR=2.5e-4	
		Diabetes pathways	DR=1.8e-3		R=7.9e-3	R=4.1e-3	FDR=7.5e-4	
	)	Cytochrome P450 – arranged by substrate type	DR=2.4e-3		R=5.9e-3	R=3.9e-3	DR=2.2e-3	
Other venohiotic metabolism		Unfolded Protein Response	DR=2.4e-3		R=6.1e-3	R=4.1e-3	FDR=5.3e-4	
		Fatty acid, triacy/glycerol, and ketone body metabolism Phase 1 – Functionalization of compounds	R=3.7e-3		DR=3.0e-3	3.3e-2	DR=1.9e-3	
		Drug metabolism - cvtochrome P450	R=5.0e-3		R=5.2e-3	(=9.0e-3	R=6.7e-3	
		Ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins	t=1.1e-2		0R=4.2e-3	R=4.3e-3	FDR=1.2e-3	
		Metabolism of lipids and lipoproteins	0e-2		R=5.9e-3	. <mark>5e-2</mark>	2e-2	
		Genes encoding structural ECM glycoproteins			FDR=1.3e-3	R=4.4e-3	FDR=2.3e-4	
	(	Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans Xenobiotics			DR=2.6e-3	R=6.0e-3	FDR=3.8e-4	
		Phase II conjugation			R=8.6e-3	R=4.4e-3	0R=3.6e-3	
Metabolism, kinase signaling	$\prec$	HIF-1-alpha transcription factor network			=2.1e-2	4.3e-2	4. <mark>0e-2</mark>	
		Integrin Signaling Pathway			= <mark>2.3e=2</mark>	=1.5e-2	t=1.1e-2	
		Jak-STAT signaling pathway			2.7e-2	=1.6e-2	2e-2	
		insulin signaling pathway Eocal adbesion			2.9e-2	0=-2	2e-2 R=8.6a-3	
		Cell surface interactions at the vascular wall			5.1e-2	3.3e-2	3.0e-2	
Call aurface recenters DDADe	5	Porphyrin and chlorophyll metabolism	FDR=5.6e-4			R=4.1e-3	DR=2.7e-3	
Cell surface receptors, PPARa	$\prec$	PPARA Activates Gene Expression	R=4.9e-3		R=4.8e-3	_	R=6.5e-3	
		semble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors	2.5e-2			.8e-2	7e-2	
		Activation of Gones by ATEA				DR=1 3e-3	FDR=7 7e-4	-1
		Glucocorticoid receptor regulatory network				5e-2	2e-2	
		Beta1 integrin cell surface interactions			DR=1.6e-3	-	DR=1.3e-3	
		Integrin cell surface interactions			DR=3.0e-3		=1.7e-2	
		Tryptophan metabolism			R <mark>=5.9e-3</mark>	FDR=5.8e-4		
		ECM-receptor interaction			= <mark>1.7e-2</mark>		R <mark>=7.8e-3</mark>	
		FGF signaling pathway Concernated to DID2 signaling in partice prices to			2.5e-2	1.6e-2		
		Pentose and olucuronate interconversions			·····		FDR=7 7e-4	
		Ascorbate and aldarate metabolism					DR=1.5e-3	
		Starch and sucrose metabolism					R=5.3e-3	
		E2F transcription factor network					R <mark>=6.5e-3</mark>	
		Metabolism of amino acids and derivatives				1.6e-2		
		Regulation of actin cytoskeleton				7e-2		
		Beta2 integrin cell surface interactions			DR=2.9e-3			
		Cholesterol biosynthesis			DR=3.0e-3			
		Integrin alphallb beta3 signaling			DR=3.0e-3			
		Platelet Aggregation (Plug Formation)			DR=3.5e-3			
		Complement and coagulation cascades			R=4.7e-3			
		Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling- Beta3 integrin cell surface integrations			N=0.98=-3			
		FOXA2 and FOXA3 transcription factor networks			=1.2e-2			
		Glutathione metabolism			=1.2e-2			
		The citric acid (TCA) cycle and respiratory electron transport			2 <mark>.5e-2</mark>			
		Cytokine Signaling in Immune system			4. <mark>7e-2</mark>			25
		Response to elevated platelet cytosolic Ca2+			.9e-2			1
		Integrins in angiogenesis: Chamaking gigagling pathway	20-2		. <mark>(e=2</mark>			
		Oremokine signaling pathway Developmental Riology	0e-2					
		Developmental biology						-≯ 📐
				Enrichmer	t significance: _lo		5 2 7 0	1
					n aigrinicarice. – It	910(101)		

3-7 ring PAH hypothesis for Petroleum Substances

The percentage weight of 3-7 ring PAHs in the UVCB is the most active contributor to the bioactivity observed



#### Cell-specific Gene Expression Signature and PAH(3-7 ring)

		1.0	- 00			HEO		
Cell type	Correlation of gene expression with PAH content	on Score	75 -					Class
A375	0.24	ressi					CGO	<ul><li>AE</li><li>FO</li><li>GO</li></ul>
iPSC CM	0.11		50 -			HFØDAE		OXASPH
iPSC ENDO	0.18	gene					HEO	BO GAS
iPSC HEP	0.75	yte (		HEC				<ul><li>WAX</li><li>D.FUEL</li><li>P.LAT</li></ul>
MCF7	0.20		25 -	GAS RAE V	индо	HEO		
<b>iPSC NEUR</b>	0.09	Hep			HERAE HEO UNTO UD SRGO UNTO UATO HEROHO UATO HERO	UDAE		
All Cells	0.47			КЕК ОLBO НО НО НЕО КЕК ОLBO СОВО НРИ Р СЕВОССКО ВТ УНСО ВЦЕОССКО ВТ УНСО ВЦЕОССКИ СТАТОТОТОТОТОТОТОТОТОТОТОТОТОТОТОТОТОТОТ	SROO UDAE			
		0.0	- 00	0.0 0.5	PAH (3-7 ring	。 g) Score		

Relationship between Hepatocyte gene expression and PAH(3-7)

Almost all of gene expression signal in Hepatocytes correlates with PAH (3-7 ring) content

PAH (3-7 ring) Score

Spearman (rank) correlation = -0.85

Principal components (PC1-3) of the "Residual" gene expression



### WP4b: Gene expression connectivity mapping based work flow





## Heatmap of C-map scores for each UVCB, pairwise ordered by PAH weight

PAH weight 10.01/100 PAH weight

Best correlation between gene-expression profiles of the UVCBs occur with those that have the highest PAH weight

Heatmap of the correlation scores ordered by the log 10 PAH[(Wt% \* (Ring 3-7 PAHs)+1]



#### Connectivity score example for HFO

Example for 034\_HFO, which is the one with the highest 3-7 ring PAH content amongst the UVCBs (top right in the heatmap, previous slide)

This sample induces the highest number of differentially expressed genes

UVCB Connectivity score vs PAH content





### Overall conclusions on transcriptomics data

- Gene expression data are useful for elucidating mechanisms behind biological responses
- High throughput gene expression profiling methods are needed to handle the large number of samples and experimental conditions that are needed for grouping
- In this instance, the gene expression data provide support that 3-7 ring PAH content is driving most of the bioactivity
  - Hepatocytes were most responsive (highest differential gene expression)
  - Highest expressed genes were involved in PAH metabolism related pathways
  - Other cell-types did not provide additional mechanistic information
- This was confirmed by the correlation between 3-7 ring PAH content with gene expression
- Connectivity mapping are a useful tool for comparing multiple gene expression profiles from multiple UVCBs
  - Gene expression profiles between substances with high PAH level are similar
- Gene expression data were overall not as informative as cell-based endpoint measurement
  - Generated transcriptomics data are still informative to add further support to the PAH hypothesis and adding a mechanistic component in combination with other data



18

# Thank you for your attention

Shu-Dong Zhang sd.zhang@ulster.ac.uk



www.concawe.eu/cat-app

