The Application of Metabolomics Platforms for Exploring Cancer Etiology

Dr. Mazda Jenab
Section of Nutrition and Metabolism
International Agency for Research on Cancer (IARC-WHO)
Lyon, France

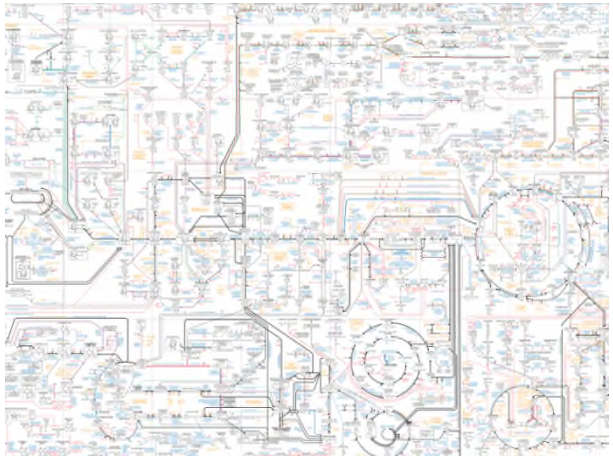
Contact email: jenabm@iarc.fr

International Agency for Research on Cancer
World Health Organization

epic
European Prospective Investigation of Cancer

What is Metabolomics?

- The study of metabolites related to specific cellular processes
- Is a powerful means of detecting disease-induced metabolic alterations
  - **The ideal metabolomic study:**
  - (i) Provides information on every metabolite in the organism, and
  - (ii) Gives insight into metabolic response to a biological situation (e.g. health / disease), exposures (e.g. diet, lifestyle, environmental, endogenous, exogenous) or experimental manipulation (e.g. intervention)
  - **Assumption 1:** every metabolite will be measured
  - **Assumption 2:** all the measurements will be biologically informative



Main Goals of a metabolomics method are to:

- obtain high coverage of the metabolome
- be fast, robust
- maximize annotation and quantification of metabolites

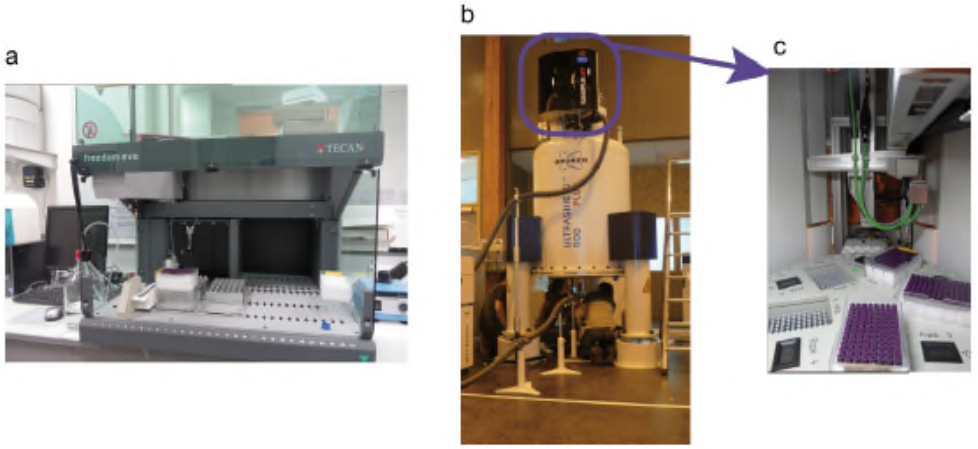
What Approaches Are Available for Metabolomics Studies?

- **Approaches:** Targeted versus Untargeted
  - **Targeted:**
    - specific compounds are quantified and compared to established reference ranges
    - methods can be developed, or commercially available kits can be used (e.g. www.biocrates.com)
    - quantification can be full or semi
    - usually, there is an inverse association between quantification and metabolite coverage (i.e. more emphasis on quantification, less range of metabolites assessed)
  - **Untargeted:**
    - analyze all detectable metabolites (known and unknown) to determine whether they are significantly different between comparison groups and then perform metabolite identification / annotation
    - is an agnostic, discovery process
    - allows data mining to identify additional compounds
    - challenges: lab-to-lab variability in methodology, data management and statistical analysis strategies

What Methods Are Available for Metabolomics Studies?

- **Methods: Nuclear Magnetic Resonance (NMR)** and Mass Spectrometry (MS)

NMR:



- High throughput, rapid
- Automated sample preparation
- Non-destructive to samples
- Can be applied to intact tissues
- Limited sensitivity, range of metabolites (~ 200: amino acids, ketone bodies, organic acids, creatinine, lipids, lipoproteins)

Ref: Markley et al., 2017, Curr Opin Biotechnol

What Methods Are Available for Metabolomics Studies?

- **Methods:** Nuclear Magnetic Resonance (NMR) and **Mass Spectrometry (MS)**

MS:

Agilent 1290 UHPLC, 6550 QTOF

- High throughput, ~600 samples / batch
- Robust methodology
- Low sample volume requirements
- High resolution
- Very wide range of features and metabolites feasible, > 6000

- Four orthogonal methods for maximum feature coverage:
  - Revered phase (RP+, RP-)
  - Hydrophilic interaction (HILIC+, HILIC-)

	Hydrophilic metabolites	Lipophilic metabolites	Cationic metabolites	Anionic metabolites
RP+		X	X	
RP-		X		X
HILIC+	X		X	
HILIC-	X			X

Annotation:

- Pure chemical standards
- In-house library, ~800
- On-line databases

METLIN

hmdb

What are the Applications for Metabolomics?

Assessment of exposure biomarkers: The Exposome

- The totality of environmental exposures received by an individual during their lifetime, and which may influence their health or the diseases that they may develop

Dietary patterns
Foods, food items
Food additives
Contaminants

Pharma-ceuticals

Pollutants

Digestion

Microbial Metabolism

Host Metabolism

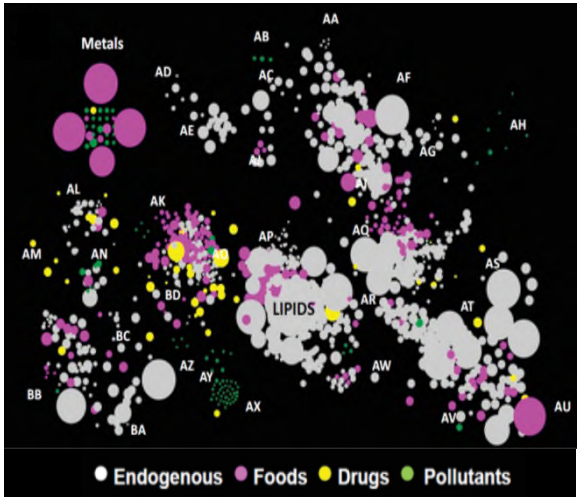
Endogenous metabolome

Food metabolome

Drug metabolome

Pollutant metabolome

Exposure Biomarkers

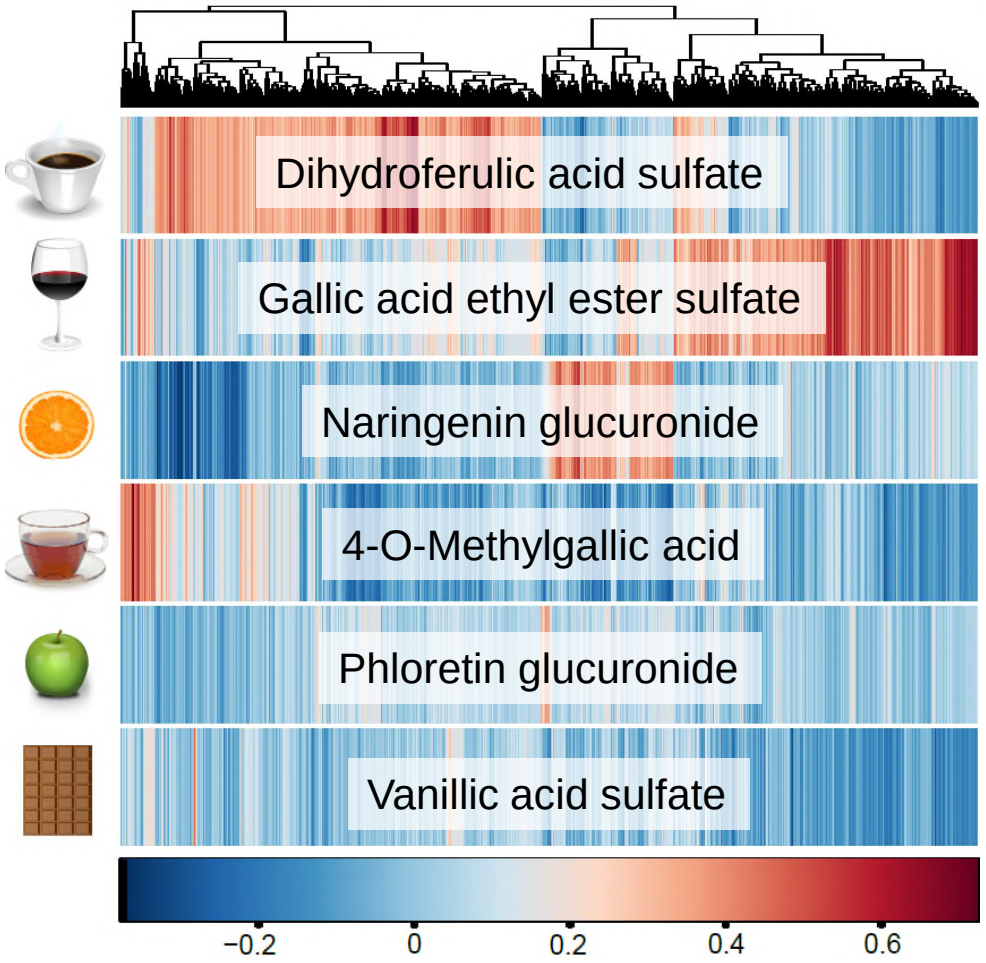


Refs: 1) Vineis P, Wild CP et al, The exposome in practice: Design of the EXPOsOMICS project. Int J Hyg Environ Health. 2017;220(2 Pt A):142-151.
2) Wild CP, Scalbert A, Herceg Z. Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. Environ Mol Mutagen. 2013 Aug;54(7):480-99

What are the Applications for Metabolomics?

Assessment of exposure biomarkers: Example of Dietary Biomarkers

- 494 subjects from 4 countries
  (France, Germany, Greece, Italy)
- 24-hr Dietary recalls & FFQ
  (434 dietary variables)
- 24-hr Urine samples
- High-resolution mass spectrometry
  (UPLC QTof)
- Iterative regression analyses
- Metabolite annotation

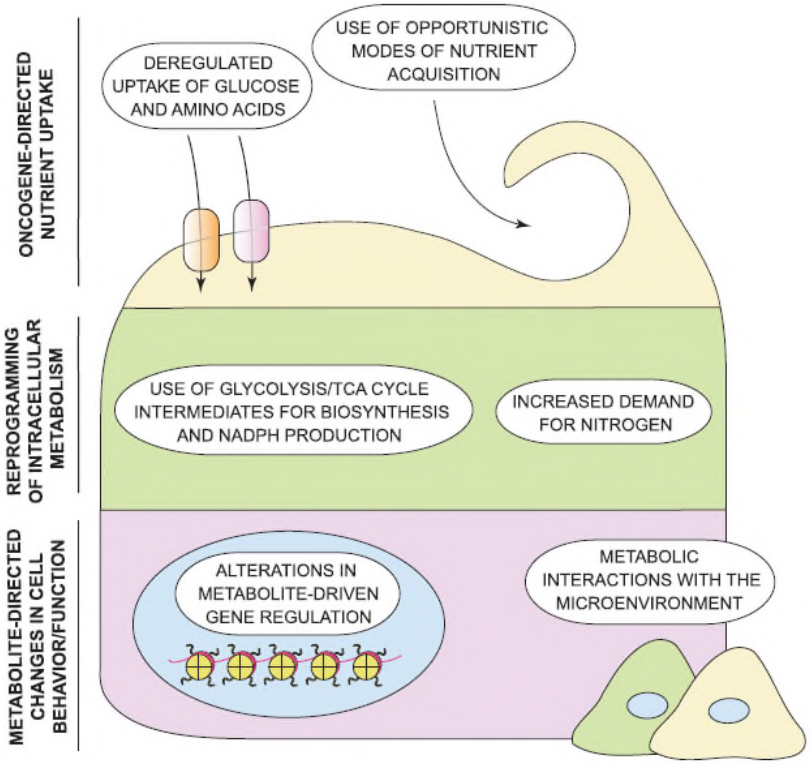


- 2,272 mass spectrometry features associated with six different foods, >80 biomarkers

Ref: Edmands et al., 2015, Am J Clin Nutr

Metabolomics in Cancer Research

Cancers are metabolic diseases:

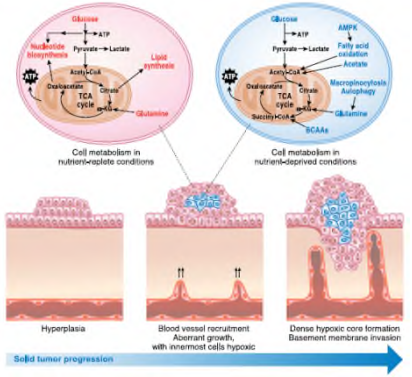


- Metabolic dysregulation is a mechanism of tumour development

Metabolic pertubations associated with cancer development may be detectable from a tumour's earliest stages

Metabolomics can provide:

- **insight into cancer etiology**, identify new risk factors
- identify biomarkers of **early diagnosis**, prognosis
- resource: prospective cohort studies with bio-samples collected prior to diagnosis



Refs: 1) Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab. 2016,12;23(1):27-47.
2) DeBerardinis RJ, Chandel NS, Fundamentals of cancer metabolism. Sci Adv. 2016, 2(5):e1600200.

Application of Metabolomics in Prospective Cohort Studies

Cancer	Reference	Cohort	Cases (controls)	Metabolome analyses	Main metabolites associated with cancer
Breast	Bro, 2015	EPIC	419 (419)	NMR	27 spectral signals
	Playdon, 2016	PLCO	621 (621)	MS	19 metabolites associated with ER+ breast cancer
Colorectum	Cross, 2014	PLCO	254 (254)	MS	Glycochenodeoxycholate (women)
	Perttula, 2016	EPIC-Italy	95 (95)	LC-MS	Long-chain hydroxylated fatty acids
Hepatocellular	Fages, 2015	EPIC	114 (122)	NMR	16 metabolites
	Stepien, 2016	EPIC	147 (147)	MS (Targeted)	various amino acids, biogenic acids and their ratios
Pancreas	Mayers, 2014	4 cohorts	453 (898)	LC-MS	Isoleucine, leucine, valine
Prostate	Mondul, 2014	ATBC	74 (74)	MS	1-Stearoylglycerol, glycerol, alpha-ketoglutarate
	Mondul, 2015	ATBC	200 (200)	MS	Inositol-1-phosphase, phospholipids, alpha-ketoglutarate and citrate, thyroxine, TMAO
	Huang, 2017	ATBC	137 (200)	MS	3-Methylhistidine, 2'-deoxyuridine, oleoyllinoleoylglycerophosphoinositol, Secondary bile acid lipids, sex steroids and caffeine-related xanthine metabolites
	Schmidt, 2017	EPIC	1077 (1077)	MS (Targeted)	Citrulline, glycerophospholipids (PC ae C30:0), acylcarnitine C3, methionine, trans-4-hydroxyproline, biogenic amine ADMA, hexose, sphingolipid SM (OH) C14:1, glycerophospholipid PC aa C42:4
Breast, prostate, colon/rectum	Kuhn, 2016	EPIC	835 (774)	MS (Targeted)	LysoPC a C18:0

The EPIC Cohort

Main Objective:

To investigate the relationship between diet, lifestyle, metabolic and genetic factors and risk of developing cancer [and other chronic diseases].

The EPIC Cohort

Main Objective:

To investigate the relationship between diet, lifestyle, metabolic and genetic factors and risk of developing cancer [and other chronic diseases].

What is it?

A large prospective cohort composed of >520,000 participants from 23 centers in 10 European countries.

Tromsø

Umeå
Malmö

Aarhus
Copenhagen

Oxford
Cambridge

Potsdam
Heidelberg

Utrecht
Bilthoven

Paris (nationwide)

Turin
Milan
Florence
Naples
Ragusa

Oviedo
San Sebastian
Pamplona
Murcia
Granada

Athens (nationwide)

Strengths of the study:

- large size
- inclusion of multiple populations
- geographic heterogeneity
- heterogeneity of dietary intakes, dietary patterns, lifestyle habits, cancer rates
- Mostly general population*

*except France (women in health insurance system or state school employees) and NL-Utrecht (women undergoing breast cancer screening)

The EPIC Cohort

Enrolment of healthy participants

Follow-up periods:
2004, 2007, 2012, 2016

Collection of questionnaire information and biological samples

Follow-up for disease status

diagnosis

vital status

Exploration of Cancer Etiology:
Dietary, lifestyle, metabolic and genetic determinants of cancer development and survival

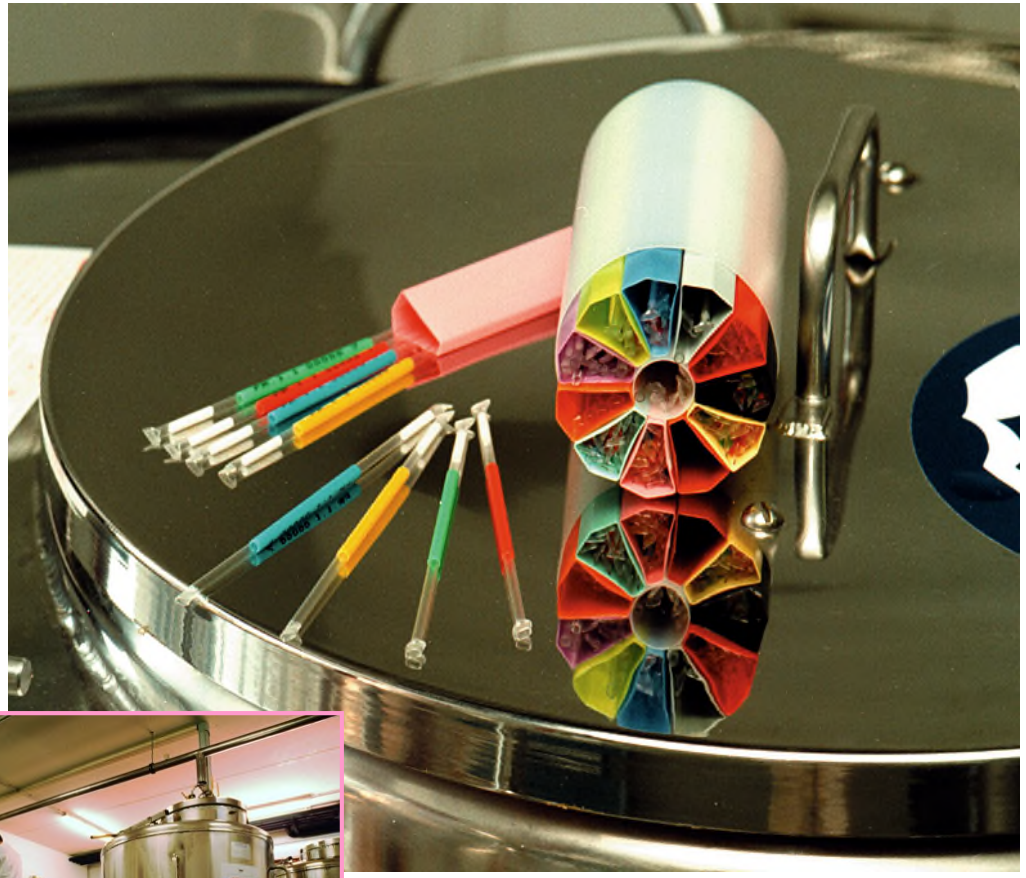
- Potential to explore markers of early diagnosis in cases diagnosed close to baseline

Pilot studies underway to assess feasibility of:

- end-point refinement by collection of tumor samples, molecular / genetic sub-typing, DNA extraction
- collection of clinical data on treat-ments, recurrence

The EPIC Biobanks

- Biological samples collected from ~80% of the cohort
- Standardized methods in all countries (except Denmark and Sweden)
- > 4.5 million aliquots collected and stored in straws at IARC, mirrored in each centre (except Denmark, Sweden)
- Straws (~450ul) of plasma (12), serum (8), buffy coat (4), red blood cells (4) stored under liquid nitrogen at -196ºC
- > 1.3 million aliquots collected and stored in Denmark and Sweden
  - Denmark: under liquid nitrogen vapour at -120ºC
  - Sweden: standard -80ºC freezers

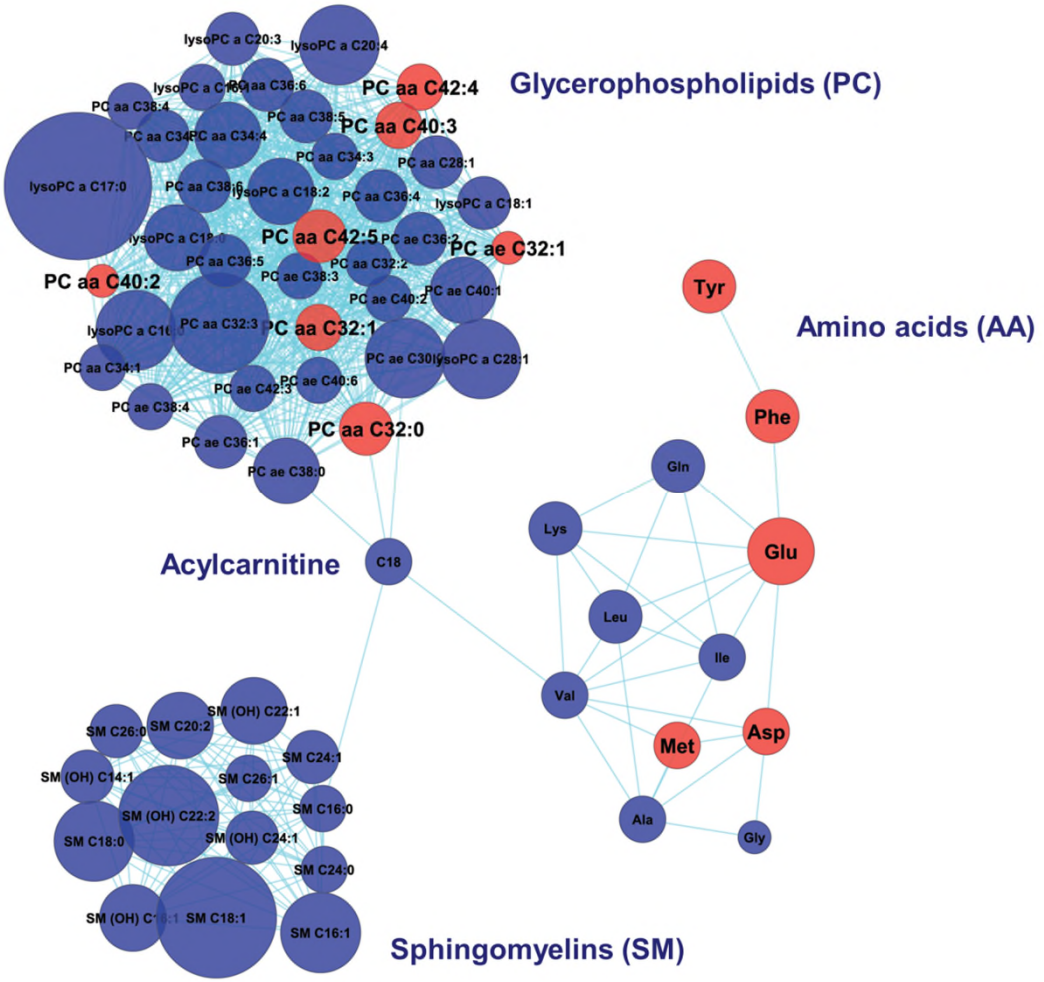


* Procedures differed in Denmark, Sweden

Targeted Metabolomics - Biocrates Kit

- Biocrates Absolute***IDQ***® p180 Kit (Biocrates Life Sciences, Austria)

Metabolite group	No. of metabolites	FIA-MS/MS	LC-MS/MS
Amino acids and biogenic amines	40		X
Acylcarnitines	40	X	
Lyso-phosphatidylcholines	14	X	
Phosphatidylcholines	74	X	
Sphingomyelins	14	X	
Hexose	1	X	
Total	183		

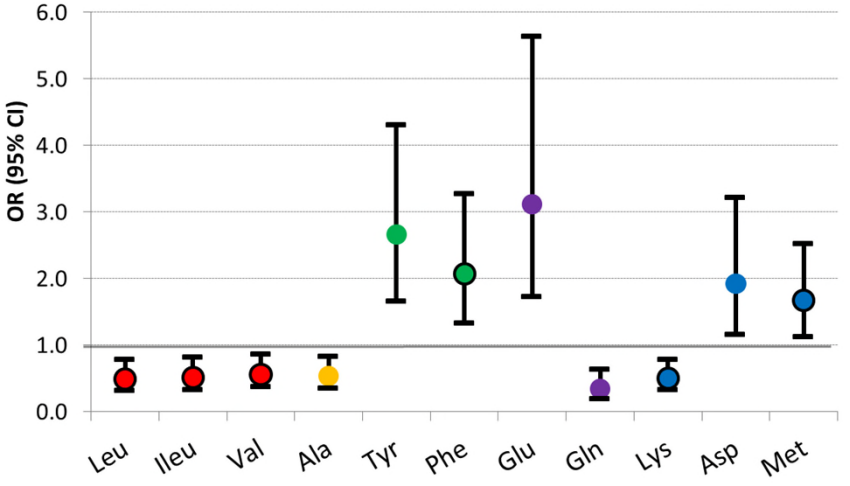


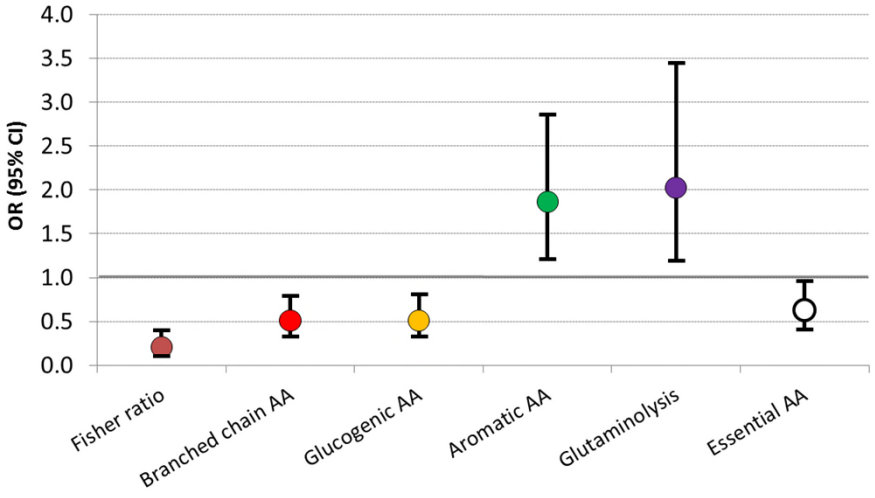
FIA-MS/MS: Flow injection analysis tandem mass spectrometry
LC-MS/MS: Liquid chromatography mass spectrometry

Connections are defined by chemical similarity score > 0.8 for lipids and 0.7 for rest of the compounds. Node size reflects multivariable adjusted OR and node colour reflects the significant decrease (**blue**) or increase (**red**) in HCC risk.

Targeted Metabolomics - HCC
Biocrates Kit – Amino Acids

- **Literature observations:**
  - NAFLD / NASH: increased blood tyrosine, glutamate,
  - Fibrosis / Cirrhosis: decreased blood branched chain AA
  - in liver disease glutamate is up-regulated while glutamine down-regulated (Beygolu & Idle, 2013): decreased glutamine synthesis and increased glutaminolysis
  - in hepatic failure circulating BCAA are decreased while aromatic AA are increased
- We observe similar changes = early disruptions in AA metabolism involved in HCC

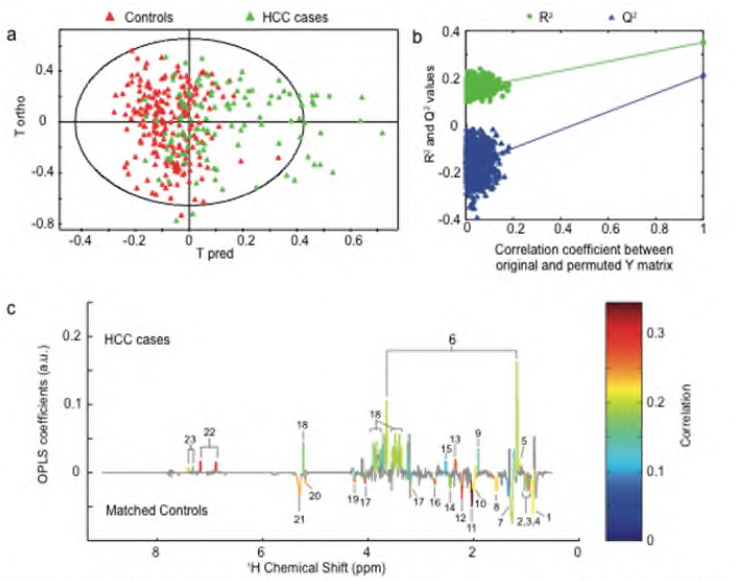




Fisher ratio= BCAA / AromaticAA (Phe + Tyr); **Branched chain AA (BCAA)**= Ileu + Leu + Val ; **Glucogenic AA**= Ale + Gly + Ser
Aromatic AA= Phe + Trp + His + Tyr ; **Glutaminolysis**=(Glu + Ala + Asp) / Gln ; Essential AA= Leu+His+Ile+Lys+Met+Phe+Thr+Trp+Val

Ref: Stepien M, Jenab M et al; Int J Cancer. 2016;138(2):348-60.

Untargeted Metabolomics - HCC Etiology
Nuclear Magnetic Resonance (NMR)



Discrimination between HCC cases (n = 114) and matched controls (n = 222). (a) O-PLS score plot of NMR spectra, $R^2(Y)$ = 0.35, Q^2 = 0.21. (b) Model validation, 1000 resampling. (c) O-PLS metabolic signature colored according to the correlation between NMR variables and case-control status after significance to ANOVA tests followed by Benjamini-Hochberg multiple correction (non-significant NMR variables are colored in grey). 1, C**H**$_3$ bond of lipids; 2, Leucine; 3, Isoleucine; 4, Valine; 5, Propylene glycol; 6, Ethanol; 7, C**H**$_2$ bond of lipids; 8, C**H**$_2$-CH$_2$-COOC bond of lipids; 9, Acetate; 10, C**H**$_2$-CH= bond of lipids; 11, N-acetyl glycoproteins; 12, Acetone and CH$_2$-C**H**$_2$-COOC bond of lipids; 13, Glutamate; 14, Glutamine; 15, citrate; 16, =CH-C**H**$_2$-CH= bond of lipids; 17, Choline; 18, Glucose; 19, Lipid O-C**H**$_2$; 20, mannose and lipids; 21, C**H**=C**H** bond of lipids; 22, Tyrosine; 23 Phenylalanine.

Crude and multivariable-adjusted OR (95% confidence intervals) of HCC risk by serum metabolites.

	Crude			Multivariable [1]		
	OR (95% CI)	*p*-value		OR (95% CI)	*p*-value	
Phenylalanine	1.83 (1.37-2.44)	4E-05	***	2.07 (1.40-3.06)	3E-04	***
Tyrosine	2.00 (1.51-2.65)	1E-06	***	2.46 (1.65-3.66)	9E-06	***
Valine	0.69 (0.54-0.88)	3E-03	**	0.82 (0.61-1.11)	2E-01	
Isoleucine	0.64 (0.49-0.83)	7E-04	***	0.72 (0.53-0.98)	4E-02	*
Leucine	0.54 (0.41-0.71)	1E-05	***	0.60 (0.43-0.85)	4E-03	**
Glutamine	0.66 (0.51-0.84)	1E-03	**	0.75 (0.54-1.03)	8E-02	
Glutamate	2.37 (1.64-3.41)	4E-06	***	2.44 (1.54-3.87)	1E-04	***
Glucose	1.53 (1.18-1.99)	2E-03	**	1.67 (1.19-2.35)	3E-03	**
N-acetyl glycoproteins	0.47 (0.36-0.62)	7E-08	***	0.46 (0.32-0.67)	3E-05	***
Citrate	1.36 (1.06-1.76)	2E-02	*	1.76 (1.22-2.54)	2E-03	**
Acetate	1.40 (1.08-1.82)	1E-02	*	1.20 (0.90-1.60)	2E-01	
Choline	0.37 (0.27-0.51)	2E-09	***	0.45 (0.31-0.65)	2E-05	***
Total lipids	0.58 (0.45-0.76)	5E-05	***	0.52 (0.36-0.74)	2E-04	***
PUFA	0.29 (0.18-0.46)	2E-07	***	0.36 (0.21-0.63)	3E-04	***
Propylene glycol	3.07 (1.58-5.97)	9E-04	***	2.20 (1.06-4.60)	4E-02	*
Ethanol	1.76 (1.13-2.74)	1E-02	*	1.36[g] (0.90-2.05)	2E-01	

* : *p*-value < 0.05, ** : *p*-value < 0.01 : *** : *p*-value < 0.001

[1] CLR models adjusted for smoking status, ethanol at recruitment, lifetime alcohol, educational status, physical activity, BMI, serum-clot contact time and waist circumference.

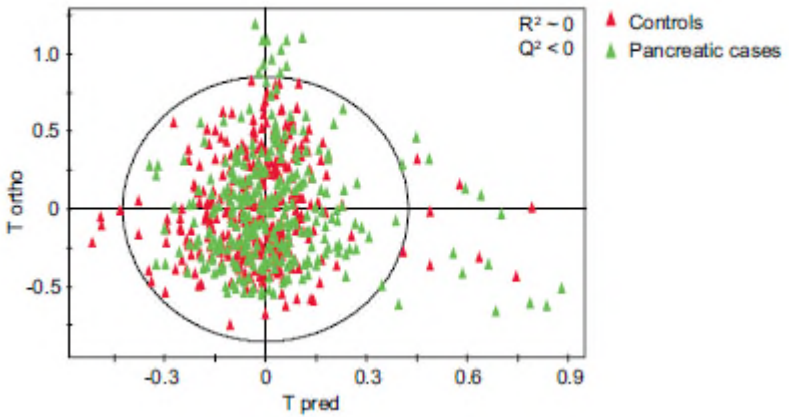
[g] OR of ethanol from multivariable CLR model are obtained after adjustment for variables mentioned for multivariable 1 expect for alcohol at baseline and lifetime alcohol.

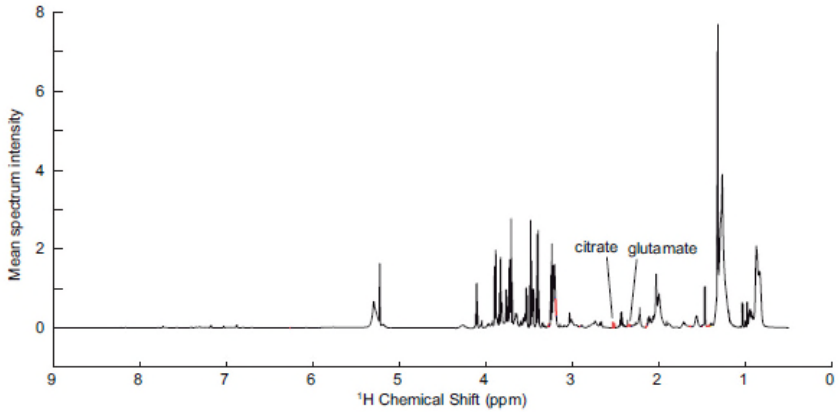
- **16 metabolites**, either endogenous or exogenous origin, found to be significantly associated with HCC risk
- Metabolic pattern associated with HCC risk comprised of perturbations in fatty acid oxidation, amino acid, lipid and carbohydrate metabolism

Ref: Fages A…and Elena B, Jenab M; BMC Med. 2015 Sep 23;13:242. doi: 10.1186/s12916-015-0462-9, PMID: 26399231

Untargeted Metabolomics – Pancreatic Cancer Etiology (NMR)

- 298 first incident pancreatic adenocarcinoma cases and 298 matched controls





Results of unadjusted and adjusted conditional logistic regression models to reveal the association of citrate and glutamate with pancreatic cancer risk.

	CLR unadjusted		Adjusted CLR	
	Odds ratio	*p*-value	Odds ratio	*p*-value
Citrate	0.83 [0.79, 0.99]	0.045 *	0.90 [0.74, 1.08]	0.27
Glutamate	0.78 [0.64, 0.96]	0.02 *	0.81 [0.64, 1.01]	0.06

Values are odds ratios with 95% confidence intervals and are expressed per standard deviation increase (* : p-value < 0.05). Adjusted CLR are controlled for diabetic status, smoking status and BMI.

- **2 metabolites** associated with pancreatic cancer risk in crude models

Pancreatic adenocarcinoma: ICDO codes C250–C259 or C25.0–C25.3 and C25.7–C25.9; Endocrine pancreatic tumours (code C25.4; histology types 8150, 8151, 8153, 8155, 8240 and 8246) excluded.

Ref: Fages A,…Elena B, Jenab M, 2018, manuscript in preparation.

Untargeted Metabolomics - HCC Etiology

- Case-control study nested within the EPIC cohort (129 cases and 129 matched-controls)
- Application of 4 different LC-MS metabolomic methods = extensive metabolite coverage
- 92 metabolites associated with HCC risk after multiple comparisons correction; 46 annotated

Top 16 HCC Risk Associations of the 46 Annotated Metabolites	LC-MS Method	Multivariable OR (95% CI) *	Speculated Role in HCC Development
Retinol	RP+	**0.27 (0.16 - 0.48)**	immune function and cell growth
Dehydroepiandrosterone Sulfate	HILIC-	**0.35 (0.22 - 0.57)**	adrenal steroid, ↑ insulin sensitivity
Glycerophosphocholine	RP+	**0.44 (0.28 - 0.71)**	phospholipid precursor
γ-carboxyethyl hydroxychroman	RP+	**0.56 (0.39 - 0.81)**	hepatic vitamin E metabolism
Creatine	RP+	**0.56 (0.37 - 0.83)**	marker of liver function
Bilirubin	RP+	**1.94 (1.22 – 3.06)**	marker of liver damage
Tyrosine	RP+	**2.04 (1.30 - 3.20)**	amino acid from meat protein
N1-Acetylspermidine	HILIC+	**2.16 (1.38 - 3.37)**	polyamine synthesis in cancer cells
Isatin	RP+	**2.28 (1.38 - 3.75)**	gut tryptophan metabolism
Benzoyl-carnitine	HILIC+	**2.74 (1.69 - 4.42)**	food additive, sugary drinks
p-Hydroxyphenyllactic acid	HILIC-	**2.77 (1.58 - 4.83)**	carcinogen, tyrosine metabolite
Sphingosine	RP+	**2.79 (1.66 - 4.71)**	↑ in NAFLD and hepatitis C infection
L,L-Cyclo(leucylprolyl)	RP+	**3.25 (1.91 - 5.53)**	marker of ↑ liver fat content
Glycochenodeoxycholic acid	RP+	**3.31 (1.99 - 5.51)**	primary bile acid, carcinogenic
Glycocholic acid	RP+	**4.07 (2.32 - 7.14)**	primary bile acid, carcinogenic
7-methylguanine	HILIC+	**6.78 (3.24 - 14.18)**	exposure to methylating agents

* statistically significant after p-value adjustment for multiple comparisons

Ref: Stepien M, Keski-Rahkonen P, Scalbert A, Jenab M; Manuscript in preparation

Untargeted Metabolomics – Pancreatic Cancer

- Case-control pilot study nested within the EPIC cohort (152 cases and 152 matched-controls)
- Application of **2** different LC-MS metabolomic methods only (RP +ve / RP –ve)

Individual Metabolites	LC-MS Method	Multivariable OR (95% CI) *	p value	FDR q Value	Bonferroni p value	Speculated Annotation
M04642	RP+	**0.66 (0.47 – 0.93)**	0.0185	0.1700	0.8500	Unknown
M00144	RP+	**1.39 (1.06 – 1.82)**	0.0156	0.1700	0.7200	Unknown
M00551	RP+	**1.39 (1.06 – 1.82)**	0.0189	0.1700	0.8700	Unknown
M00720	RP+	**1.46 (1.08 – 1.97)**	0.0143	0.1700	0.6600	Unknown
M00295	RP+	**1.54 (1.05 – 2.26)**	0.0276	0.2100	1.0000	Pentose sugar
M01036	RP+	**1.73 (1.04 – 2.89)**	0.0360	0.2400	1.0000	Unknown
M02672	RP+	**1.76 (1.10 – 2.81)**	0.0184	0.1700	0.8500	Unknown
M00781	RP-	**1.46 (1.06 – 2.02)**	0.0207	0.3400	1.0000	Docosapentaenoic acid
M00205	RP-	**1.60 (1.11 – 2.30)**	0.0113	0.3400	0.5700	Unknown
M00175	RP-	**1.66 (1.11 – 2.48)**	0.0144	0.3400	0.7200	Unknown

	RP+	RP-
Total features in initial dataset	5277	2444
Features in ≥75% of samples	2270	1341
Features with > median 1.2 fold difference between cases and controls	46	50

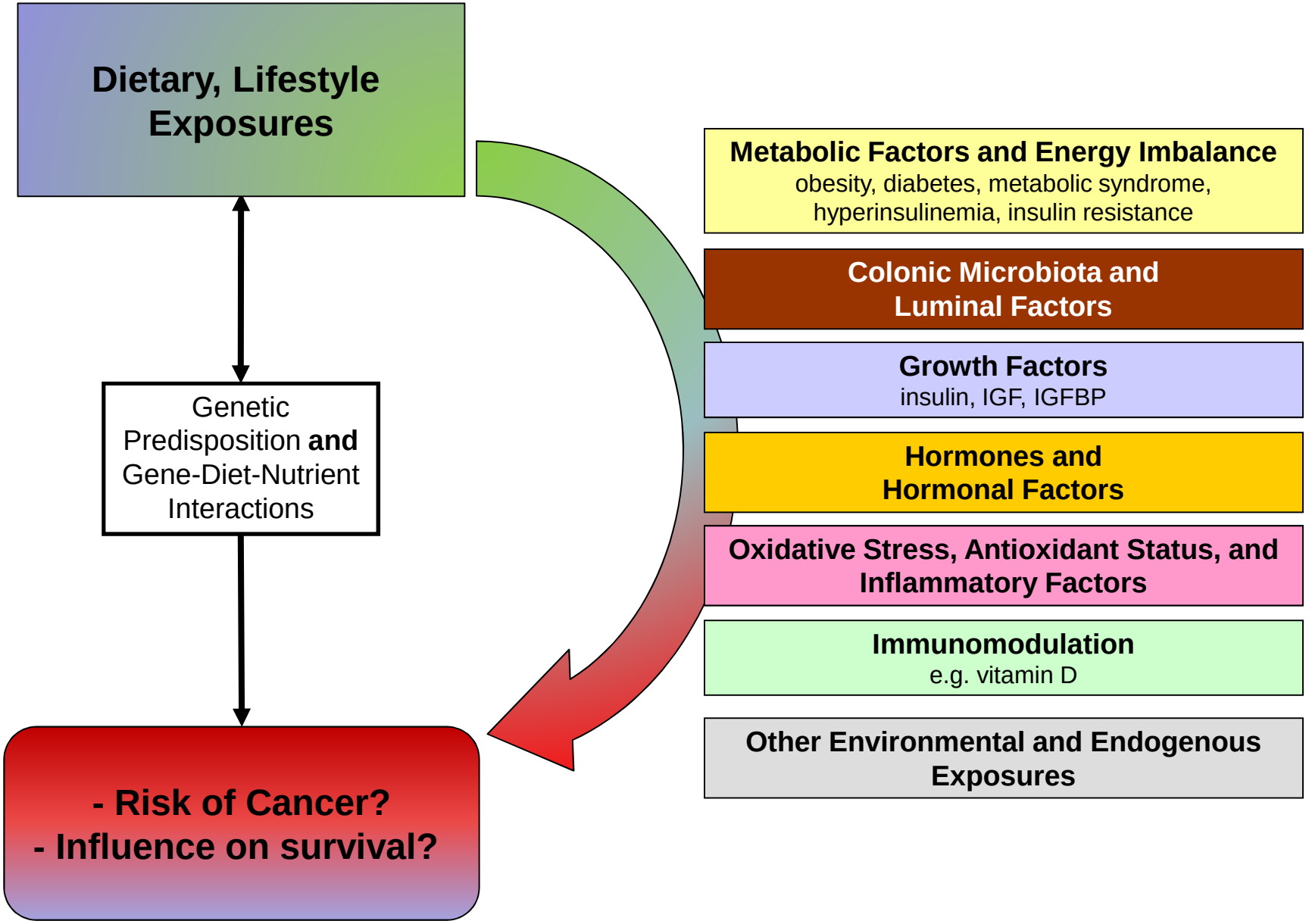
- Two further methods (HILIC +ve/ HILIC –ve) and detailed annotation still to be performed
- Expansion to entire series of pancreatic cancer cases and matched controls with available bio-sample, n=626

* statistically significant after p-value **unadjusted** for multiple comparisons.

- N cases = 152, 60 men, 92 women; average age = 57.8

epic
European Prospective Investigation of Cancer

Cancer Etiology:
Multi-factorial, Multi-mechanistic



Summary and Conclusions

- Applications of metabolomics methods can provide novel and meaningful insights into cancer etiology
  - proven utility in identifying novel exposure biomarkers
  - useful for search for biomarkers of early diagnosis
- Several approaches and methodologies exist
  - each with its own advantages, disadvantages
- Access to specialized labs is essential
  - sample treatment, annotation, interpretation are very important
- Study design, sample treatment / preparation, statistical analysis methods are all very important considerations

Acknowledgements

- **The EPIC Network:**
  - **IARC:** Marc Gunter
  - **Imperial College London:** Elio Riboli
  - **Denmark:** Anne Tjonneland, Kim Overvad
  - **France:** Francoise Clavel-Chapelon, Marie Christine Boutron-Ruault
  - **Germany:** Rudolf Kaaks, Heiner Boeing
  - **Greece:** Antonia Trichopoulou
  - **Italy:** Franco Berrino, Paolo Vineis, Domenico Palli, Vittorio Krogh, Rosario Tumino, Salvatore Panico
  - **Netherlands:** Petra Peeters, Monique Verschuren, Bas Bueno-de-Mesquita
  - **Norway:** Elizabete Weiderpass, Eiliv Lund,
  - **Spain:** Antonio Agudo, Carmen Navarro, Aurelio Barricarte, Miren Dorronsoro
  - **Sweden:** Goran Hallmans, Mattias Johansson, Jonas Manjer
  - **United Kingdom:** Tim Key, Kay-Tee Khaw, Nicholas Wareham, John Danesh
  - All EPIC-associated colleagues
- **IARC Colleagues:**
  - Augustin Scalbert (Head, IARC Biomarkers Group), Magdalena Stepien, Pekka Keski-Rahkonen, Agneta Kiss, Pietro Ferrari, Neil Murphy, David Achaintre
- **Funding Agencies:**
  - INCA (French National Cancer Institute)

CONTACT email: jenabm@iarc.fr

Institut National du Cancer

epic
European Prospective Investigation of Cancer

International Agency for Research on Cancer
World Health Organization