


Introduction to -omics and Biomarkers

Antonia Vlahou Ph.D
COST Summer School
Spetses, September 2019



Perception of reality
an evolutionary process



Outline

- General Concepts on Biomarkers
Definitions, Principles of biomarker discovery, validation, implementation
- Biomarkers in ‘systems biology’ era
Biomarker profiles
- Examples from multi-omics applications

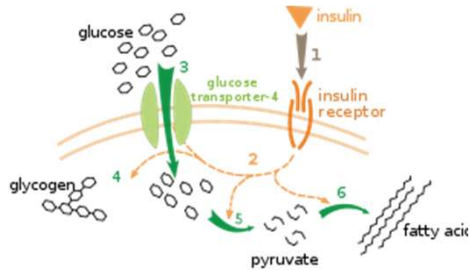
DEFINITION

“ a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention “

(National Institutes of Health Biomarkers Definitions Working Group, 1998)

“biomarker is a characteristic...”

Marker can be any biological molecule...
(genes, proteins, metabolites...)



Additional known markers: PSA, CA125, CRP ...

“objectively measured”
Need for a robust assay

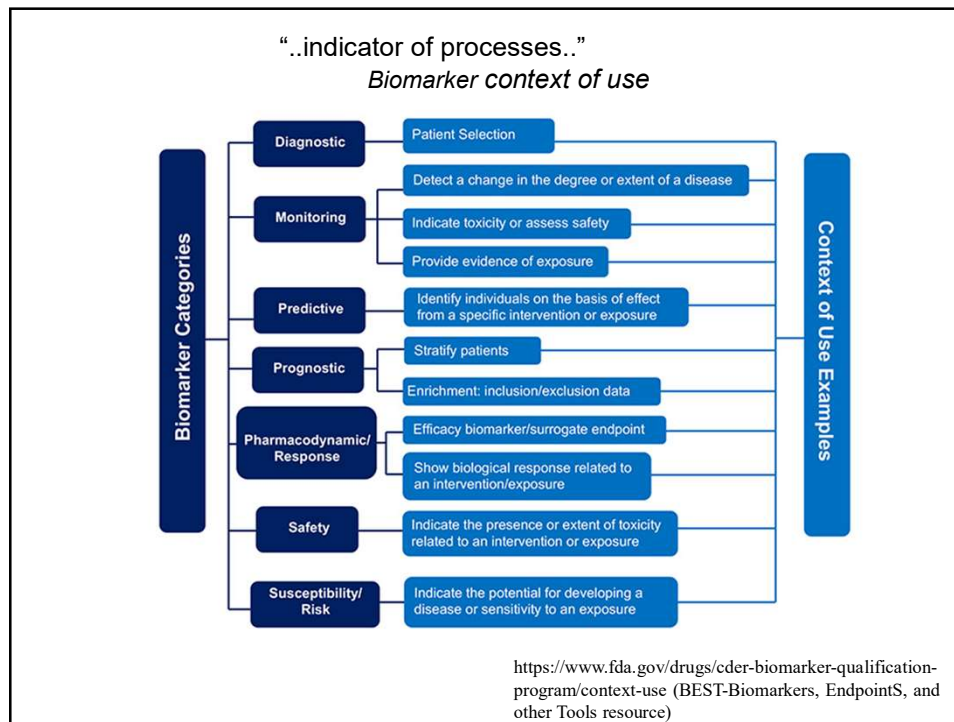
The objective of the analytical procedure should be clearly understood since this will govern the validation characteristics which need to be evaluated. Typical validation characteristics which should be considered are listed below:

- Accuracy
- Precision
 - Repeatability
 - Intermediate Precision
- Specificity
- Detection Limit
- Quantitation Limit
- Linearity
- Range

Parameters	Validation Recommendations		In-Study Analysis Recommendations
	Chromatographic Assays (CCs)	Ligand Binding Assays (LBAs)	
Accuracy and Precision (A & P)	Elements: <ul style="list-style-type: none">• A & P should be established with at least three independent A & P runs, four QC levels per run (LLOQ, L, M, H QC), and \geq five replicates per QC level.	Elements: <ul style="list-style-type: none">• A & P should be established with at least six independent A & P runs, five QC levels per run (LLOQ, L, M, H, ULOQ QC), and \geq three replicates per QC level.	Elements: <ul style="list-style-type: none">• Not applicable
	A & P Run Acceptance Criteria: <ul style="list-style-type: none">• The run should meet the calibration curve acceptance criteria and include the LLOQ calibrator.• This run has no QC acceptance criteria.	A & P Run Acceptance Criteria: <ul style="list-style-type: none">• The run should meet the calibration acceptance criteria and include the LLOQ calibrator.• This run has no QC acceptance criteria.	
	Accuracy: Within-run and between runs: <ul style="list-style-type: none">• $\pm 15\%$ of nominal concentrations, except $\pm 20\%$ at LLOQ.	Accuracy: Within-run and between runs: <ul style="list-style-type: none">• $\pm 20\%$ of nominal concentrations, except $\pm 25\%$ at LLOQ, ULOQ	Accuracy: Between runs: <ul style="list-style-type: none">• CC: $\pm 15\%$ of nominal concentrations• LBA: $\pm 20\%$ of nominal concentrations
	Precision: Within-run and between runs: <ul style="list-style-type: none">• $\pm 15\%$ CV, except $\pm 20\%$ CV at LLOQ	Precision: Within-run and between runs: <ul style="list-style-type: none">• $\pm 20\%$ CV, except $\pm 25\%$ at LLOQ, ULOQ	Precision: Between runs: <ul style="list-style-type: none">• CC: $\pm 15\%$ CV• LBA: $\pm 20\%$ CV
	Total Error: <ul style="list-style-type: none">• Not applicable	Total Error: <ul style="list-style-type: none">• QC should be $\pm 30\%$, except at LLOQ, ULOQ $\pm 40\%$	Total Error: <ul style="list-style-type: none">• Not applicable

<https://www.fda.gov/drugs/guidancecompliancer egulatoryinformation/guidances/ucm265700.htm>

<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry>



“..indicator of processes..”

Biomarker context of use

Expected clinical impact of biomarker

- Added value of marker to current diagnostic/prognostic means
 - Increase accuracy (work as adjunct)
 - Replace an invasive procedure by a non invasive test
- Marker should confer a tangible impact to disease management
 - Guide patient management and treatment



Application of biomarkers together with therapeutics

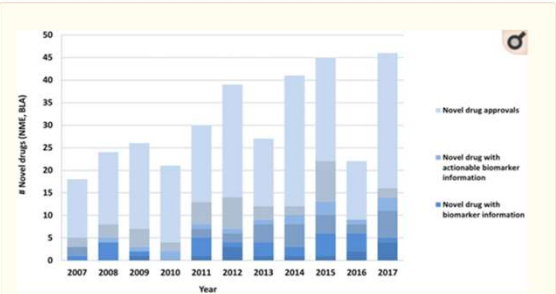


Figure 1

Novel Drugs Approved (NME/BLA) Between 2007 and 2017 With Genomic and Other Selected Biomarker Information in Labeling

Shaded areas depict oncology drugs. Actionable biomarker information refers to a specific prescribing recommendation that is included in one of the following label sections: 1) Boxed Warning, 2) Indications and Usage, 3) Dosage and Administration, 4) Contraindications, or 5) Warnings and Precautions. Biomarkers may be any genomic biomarker or other selected protein biomarker that are used for patient selection. BLA = Biologic License Application; NME = New Molecular Entity.

Drozda et al JACC: VOL. 3, NO. 4, 2018

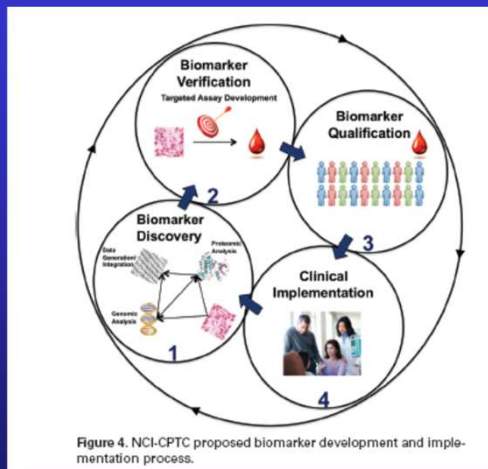
Application of biomarkers together with therapeutics

TABLE 1 Selected Examples of Nononcology Drugs With Biomarker Changes Implemented Based on the Post-Marketing Evidence					
Drug (Year Approved/Initial Pgx Revision)?	Therapeutic Area	Biomarkers	Outcome	Source of Evidence	Current Actions?
Abacavir (1998/2008)	Infectious diseases	HLA-B*57:01	Hypersensitivity reactions	Randomized controlled trial	Recommends HLA-B*57:01 testing before initiating treatment and avoiding use in HLA-B*57:01 carriers
Carbamazepine (1968/2007, 1968/2013)	Neurology	HLA-B*15:02, HLA-A*31:01	Severe cutaneous adverse reactions (e.g., SJS, TEN)	Case-control studies, meta-analysis	Recommends HLA-B*15:02 testing before initiating treatment in patients of Asian ancestry and warns prescribers about increased risk of developing hypersensitivity reactions in the presence of HLA-A*31:01
Citalopram (1998/2011)	Psychiatry	CYP2C19	QT prolongation	PD studies	Recommends a maximum dose to be used in individuals who are CYP2C19 PMs based on QT prolongation effect
Clopidogrel (1997/2010)	Cardiology	CYP2C19	Diminished antiplatelet response	PK and PD studies, retrospective case-control and cohort studies	Warns prescribers of the risk for diminished response in CYP2C19 PMs and provides consideration for use of alternate treatments in "Boxed Warning" section
Codeine (1950/2009)	Anesthesiology	CYP2D6	Respiratory depression, death	PK studies, case-series	Contraindicated in children <12 yrs of age and in children <18 yrs of age after tonsillectomy and/or adenoidectomy based on risk of respiratory depression and death in CYP2D6 UMs
Fluoxetine (1984/2011)	Psychiatry	CYP2D6	QT prolongation, sudden cardiac death	PK studies	Recommends testing at a certain dose threshold that is not to be exceeded in CYP2D6 PMs and longer dose titration interval in CYP2D6 PMs
Rosuvastatin (2003/2012)	Endocrinology	SLCO1B1	PK information	PK studies	Provides PK information in "Clinical Pharmacology" section; no alternative treatment strategies recommended based on SLCO1B1 genotype
Tramadol (1995/1999)	Anesthesiology	CYP2D6	Respiratory depression, death	PK studies, case-series	Contraindicated in children <12 yrs of age and in children <18 yrs of age after tonsillectomy and/or adenoidectomy based on risk of respiratory depression and death in CYP2D6 UMs
Valproic acid (1978/2015)	Neurology	POLG	Fatal hepatic failure	Case-series	Contraindicated in children with mitochondrial disorders resulting from POLG mutations based on risk of fatal hepatic failure

†Drug labeling was further modified over years. Information included in the "Current Actions" column was based on the information available on the Drugs@FDA website. For additional examples of drugs with biomarker information included in the labeling go to <https://www.fda.gov/drugs/developmentresearch/ucm332698.htm>. CYP = cytochrome P450; G6PD = glucose-6-phosphate dehydrogenase deficiency; PD = pharmacodynamic; Pgx = pharmacogenetic; PK = pharmacokinetic; PMs = poor metabolizers; POLG = desaminase; SJS = Stevens-Johnson Syndrome; TEN = Toxic Epidermal Necrolysis; UMs = ultra-rapid metabolizers.

Drozda et al JACC: VOL. 3, NO. 4, 2018

Steps in Biomarker Development

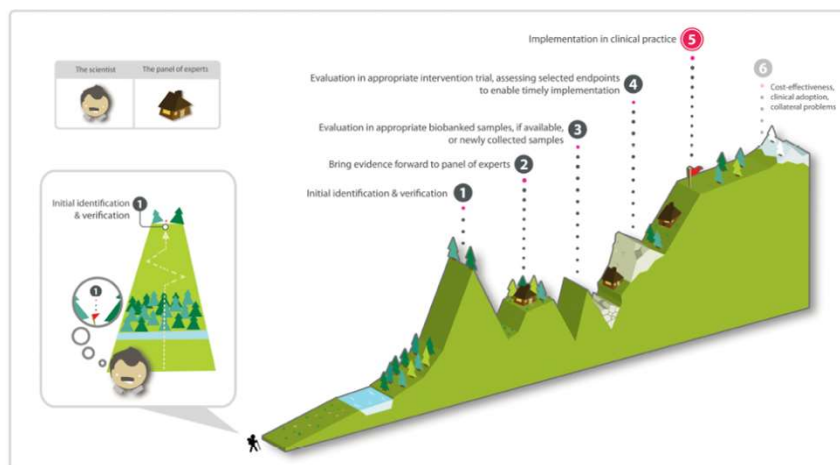


Biomarker verification-qualification

- Vast need for clinical resources
- Assay development
- High cost
- Regulatory approval

Boja et al, Journal of Proteome Research 2011, 10, 66–84

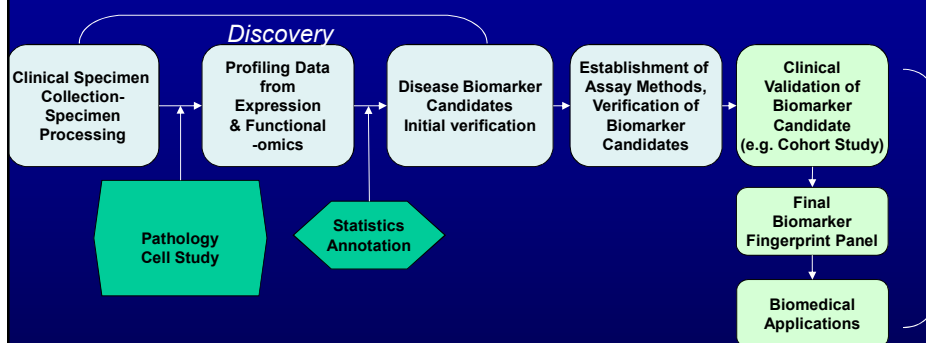
To achieve the clinical impact we need multi-disciplinary collaborations



Eur J Clin Invest. 2012 Sep;42(9):1027-36

Critical Steps in -omics analysis during biomarker discovery

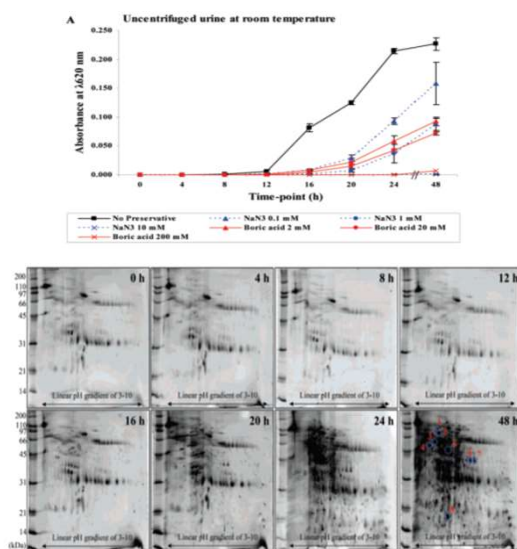
Pipeline in biomarker discovery



Study design: Clinical aspect

- Define objective keeping clinical relevance in mind
- Selection of cases- controls in the context of specific clinical question
- Often the use of healthy controls is not clinically relevant
- If targeting screening markers: compare disease to healthy controls
If targeting Biomarkers for staging:
comparison should be conducted between patients of different disease stages

Sample collection affects molecular profile



Thongboonkerd and Saetum, J Proteome Res.
2007;6(11):4173-81

Standard Operating Procedures Exist

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

Early Detection Research Network Biomarkers: the key to early detection

DCP Division of Cancer Prevention

Home About EDRN Biomarkers Protocols Science Data Publications Resources Specimens

You are here: Home / Resources / EDRN Standard Operating Procedures (SOP) / Standard Operating Procedures

Standard Operating Procedures

Announcement 01/29/2019

Blood-derived Specimen Collection Protocols

- Serum SOP
- Plasma SOP (EDTA)

Please visit the EDRN Registration page for registration/holder information for the 34th EDRN Steering Committee Meeting from March 18-20, 2019, in Nashville, TN.

Urine Collection Protocols

- EDRN Urine Collection Project (for urine reference sets)

Announcement

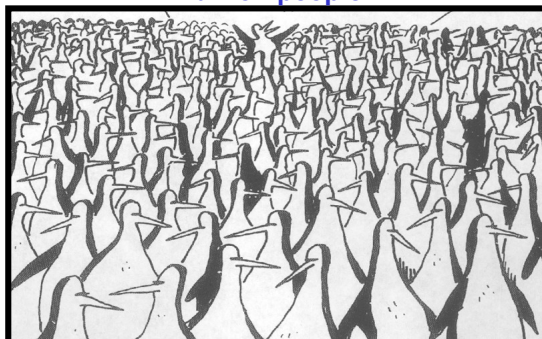
Increased complexity generates technical difficulties

Large dynamic range (especially for proteomics-metabolomics)

Dynamic range of serum proteins

50 mg/ml	Albumin
1 mg/ml	Haptoglobulin
1 µg/ml	IgE
1 ng/ml	CEA (10 ng/ml) CK-MB (3 ng/ml) PSA (1 ng/ml)
1 pg/ml	Troponin T (10 pg/ml) TSH NT-pro-BNP

... situation is similar to trying to spot one single face amongst 4 billion people...



No technique is perfect!!!

Each approach has its own strengths and limitations!!

Approach	Major Strengths	Major weaknesses	Application (marker development)	Other characteristics
2DE- MS	Retains Information at protein level	Tedious-low throughput and resolution	Discovery	Relative quantification
LC-MS (shot-gun)	High resolution	Tedious-Low throughput	Discovery	Complexity in data analysis-Currently most widely employed technique for marker discovery
LC-MS (MRM)	High throughput – absolute quantification	Proteotypic peptides required- High cost related to labeling	Validation	Promising alternative to Elisa for targeted quantification
CE-MS	High resolution for native peptidome-high reproducibility	Peptide identification requires change in platform	Discovery- Validation- Implementation	Established assay for urinary peptide profiling-applications to running clinical trial for CKD

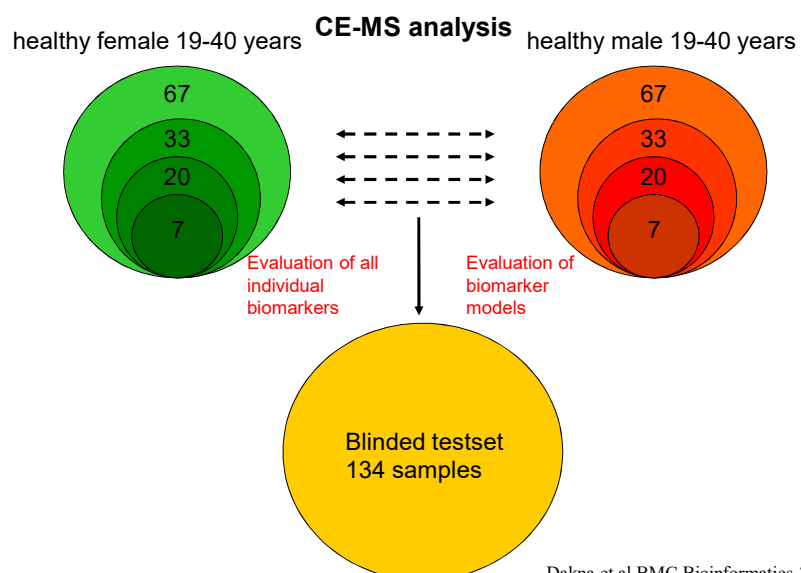
Statistical Design is important!!

How to identify valid biomarkers and biomarker models?

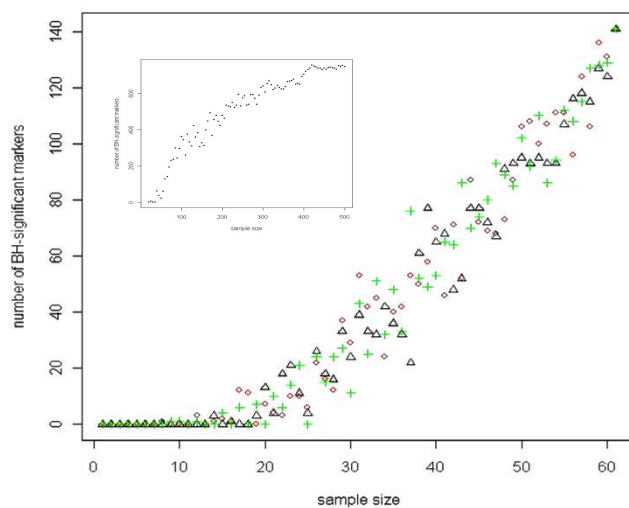
- How many samples are required to define useful biomarkers?
- Relevance of Statistics?
- Which classifiers perform best?
- Is assessment in a blinded set necessary/helpful?



Defining the value of statistics

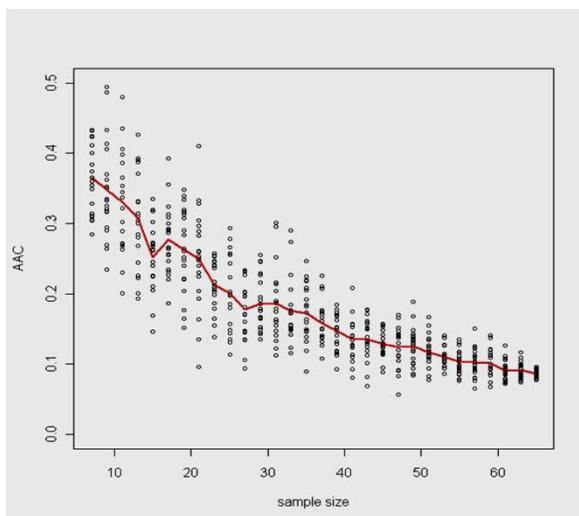


Increasing sample size results in increasing number of biomarkers

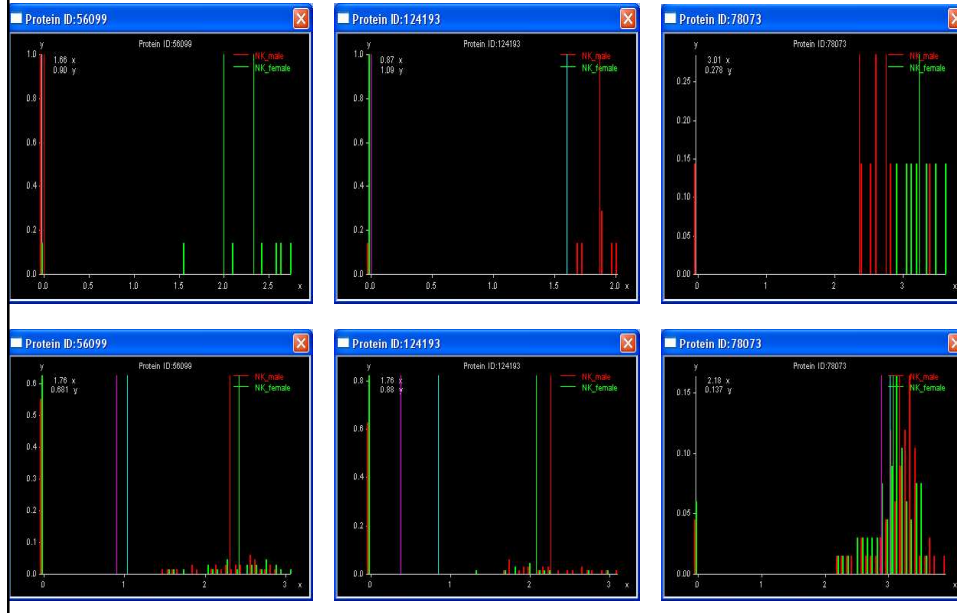


Dakna et al BMC Bioinformatics 2010;11:594.

Classification error depends on sample size



Inappropriately low numbers of independent samples result in 'erroneous biomarkers'



Data analysis: Statistics matter !

Number of significant markers depends on statistical test

	Test	No point-mass	Dissonant	Consonant	Total	Validated	% Validated
unadjusted p-values	t-test	3	0	245	248	63	25 %
	Wilcoxon	4	5	314	323	109	33 %
	Two-part-t	3	8	229	240	68	28 %
	Two-part-W	4	11	286	301	104	34 %
	Empirical LRT	4	7	271	282	81	28 %
BH-adjusted p-values	t-test	0	0	57	57	27	47.3 %
	Wilcoxon	3	1	137	141	58	41.1 %
	Two-part-t	0	3	66	69	30	43.4 %
	Two-part-W	3	6	103	112	55	49.1 %
	Empirical LRT	2	5	109	116	43	37.0 %

The multiple testing problem

- If you target a nominal p-value of 0.05, in 100 independent experiments you will receive 5 false positive results.
- In an experiment testing 2000 hypotheses (proteins), consequently $5 \times 20 = 100$ false positives ('potential biomarkers') will be returned.

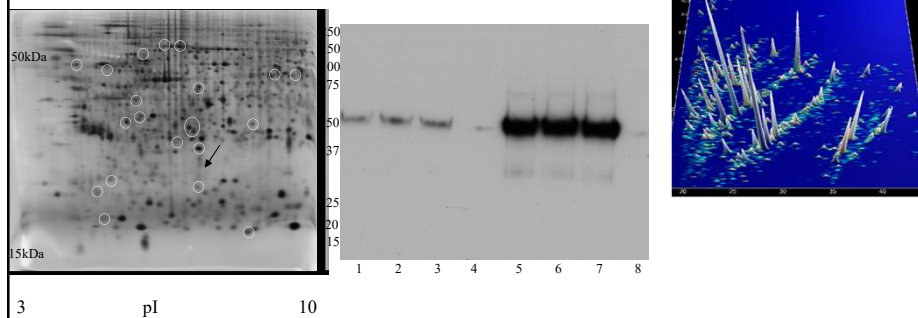
Some general principles of statistical analysis of -omics data

- N= 1 per group is not acceptable!
- Perform power calculations
- The more heterogeneous the higher the sample size: Perform Matching!
cell lines- animal models: min n=5 per group; human specimens: match!!
(min n=30 per group)
- Generally use of non parametric tests is recommended
- Perform adjustment for multiple testing
- Always validate results in a new set of samples

Data have to be confirmed in an independent test set of samples !!!

Application of alternative platform

Application of same platform



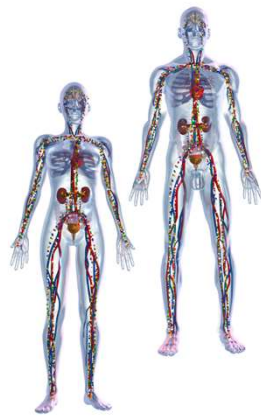
Data Interpretation !!!!!

"..biomarkers were highly specific.."

"..promising and interesting diagnostic tool..."

"....significant biomarker potential..."

Single Markers are insufficient to display disease



A complex organism, experiencing a multitude of complex environmental impacts, can not be described in detail by single features

Phenotype: a result of multi-source impacts

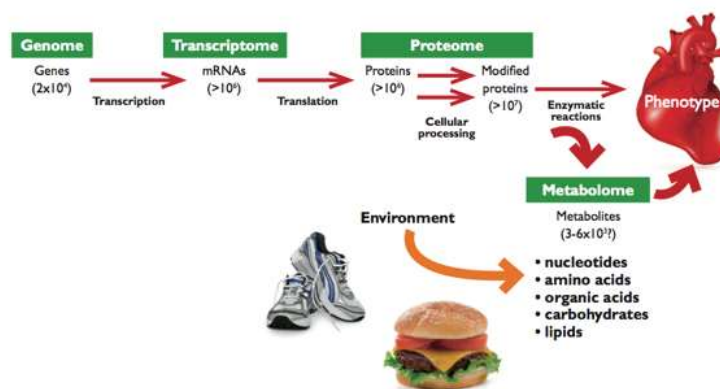
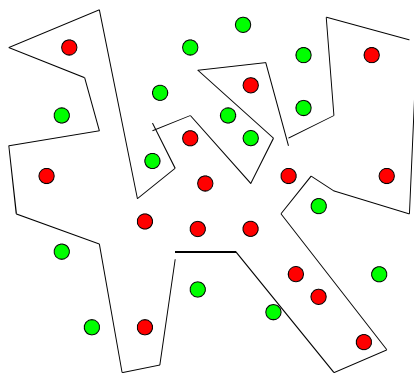


Image from harvardmagazine.com

Classifying: the curse of overfitting



Data Interpretation !!!!!

“..biomarkers were highly specific..”
 ..” promising and interesting diagnostic tool...”
 “....significant biomarker potential...”

Place findings in the context of existing knowledge!!!

- Functional annotation of proteins
- Analysis of subcellular localisation
- Literature Mining
- Pathway analysis-biological role

- Annotation of compounds
- Functional Annotation of targeted/ affected proteins
- Literature mining, Patent Search

BIOMARKERS

Recommendations for Biomarker Identification and Qualification in Clinical Proteomics

Harald Mischak,*† Günter Allmaier, Rolf Apweiler, Teresa Attwood, Marc Baumann, Ariela Benigni, Samuel E. Bennett, Rainer Bischoff, Erik Bongcam-Rudloff, Giovambattista Capasso, Joshua J. Coon, Patrick D'Haese, Anna F. Dominiczak, Mohammed Dakna, Hassan Dihazli, Jochen H. Ehrlich, Patricia Fernandez-Llama, Danilo Fliser, Jorgen Frokiaer, Jerome Garin, Mark Girolami, William S. Hancock, Marion Haubitz, Denis Hochstrasser, Rury R. Holman, John P. A. Ioannidis, Joachim Jankowski, Bruce A. Julian, Jon B. Klein, Walter Kolch, Theo Luiders, Ziad Massy, William B. Mates, Franck Molina, Bernard Monsarrat, Jan Novak, Karlheinz Peter, Peter Rossing, Marta Sánchez-Carbayo, Joost P. Schanstra, O. John Semmes, Goce Spasovski, Dan Theodorescu, Visith Thongboonkerd, Raymond Vanholder, Timothy D. Veenstra, Eva Weissinger, Tadashi Yamamoto, Antonia Vlahou

Published 25 August 2010, Volume 2 Issue 46 46pp42

Recommended steps for clinical proteomics

- 1) Define a clear clinical question
- 2) Define the type of samples needed
- 3) Define and validate the analytical platforms
- 4) Perform study of the training set
- 5) Evaluate findings on blinded samples
- 6) Demonstrate significant improvement in clinical study
- 7) Apply towards clinical use

PERSPECTIVE

Table 1. Requirements for scientific reporting of proteomic biomarker data.

Describe and justify the clinical question, outcomes, and selection of subjects	Describe the clinical question and justify why it is of interest; describe what outcomes are assessed and comment on their clinical validity, potential for misclassification, and verification bias, if pertinent; clarify what are the eligibility criteria for the selected study populations and justify specific choices.
Describe the assessed subjects	Provide demographic information with gender, age, ethnic origin, and concomitant medications at a minimum, and all relevant disease-related and clinical parameters.
Describe sampling	Provide an accurate description of the sampling conditions and procedures (including the collection process and any manipulation of the sample before storage, the time between sampling and storage, storage conditions, and the addition of any protease inhibitors and/or preservatives). Justify the sampling choices according to the literature or supporting experimental data.
Describe experimental methodology	The procedure, as well as the observed standard deviation of technical specifications related to the procedure, should be given. To attribute the same identity to a certain feature in several independent analyses, accepted deviations of mass and other parameters (retention time, migration, position on gel, etc.) must be reported. Also, the observed deviation in identifying parameters and (relative) abundance, when the same sample is analyzed repeatedly, must be reported.
Describe the statistical evaluation	Provide details on determination of sample size, statistical analysis plan (for appraising calibration, discrimination, and/or reclassification), any consideration or adjustment for covariates (including treatment, whenever pertinent), methods for adjustment for multiplicity, and parameters used in complex machine-learning approaches, whenever pertinent. Clarify which analyses are pre-defined and which are post hoc.
Validate results	The results must be confirmed in at least one independent sample set. The sampling and characteristics of the validation population should be reported, and the analysis should be symmetrical in the test and validation data sets; any deviations should be reported.
Acknowledge limitations	No study is perfect; limitations should be clearly acknowledged and their potential impact on the results discussed.
Take responsibility	The contributions of each author should be clearly stated.

Guidelines in J of Proteomics:

Mandatory requirements for reporting of biomarker studies:

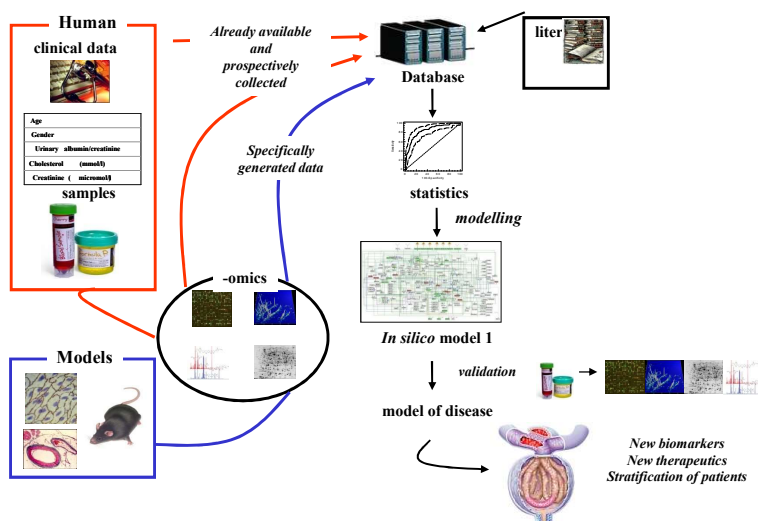
- 1) A clinical biomarker is only relevant in specific contexts of use per disease, it must have a demonstrated potential to improve the current state of the art, (either being of added value, or based on its sole performance) and its application must be linked to a clear clinical (therapeutic) consequence. As such, the specific proposed context of use of the presented biomarker must be clearly provided and the expected practical consequence of the biomarker application be discussed.
- 2) A biomarker can only be assessed in an independent (ideally blinded) test set, containing sufficient samples to demonstrate significant value and justify relevant claims regarding biomarker use. Assessment of performance in a discovery set is inappropriate.
- 3) This initial independent validation and performance assessment has to be performed in samples that reflect the typical clinical situation depending on the targeted context of use.

J Proteomics. 2014 Jan 16;96:A1-3.

Examples of Applications of multi -omics

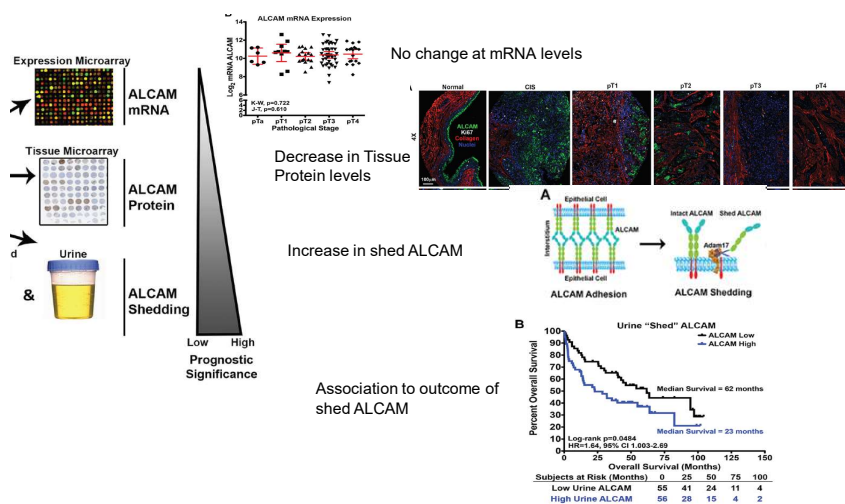


A “systems” biology approach in biomarker and drug target discovery



Modified from Nephrol Dial Transplant. 2010;25(4):1015-8

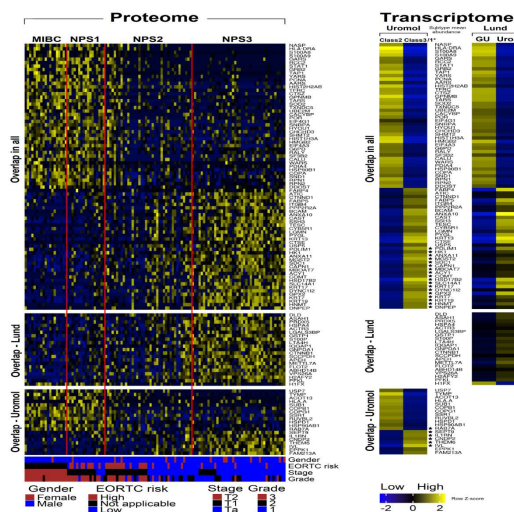
Cross-omics to understand biology *Activated Leukocyte Cell Adhesion Molecule*



Egloff et al Oncotarget, 2017, 8, (1): 722-741

Shortlisting features of prognostic potential for bladder cancer through tissue cross-omics analysis

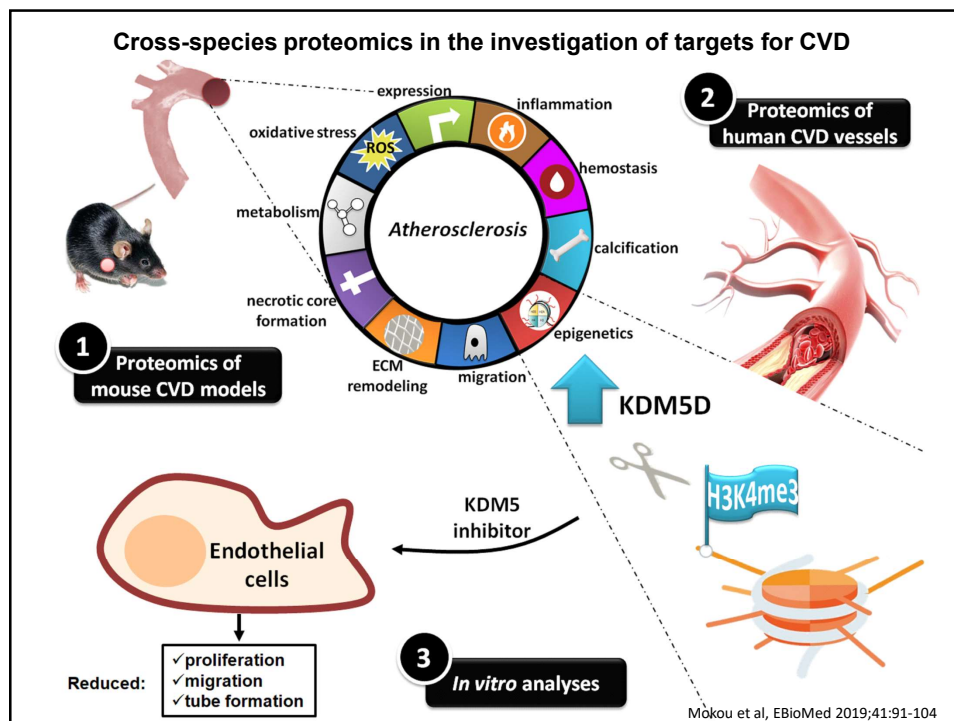
- Proteomic analysis of 119 NMIBC tissue samples
- Unsupervised clustering
- Cross-omics analysis
- Shortlisting of profiles of prognostic potential



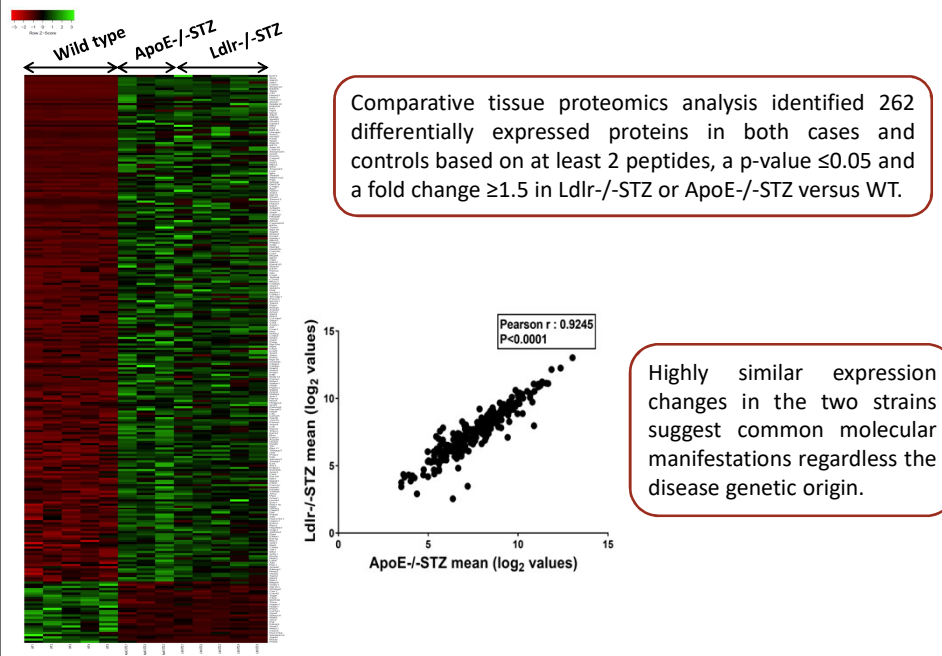
Linking tissue findings to body fluids

Evaluate protein detectability in urine and assay availability

Gene name	Description	Proteomics		UROMOL	LUND	detected in urine (in at least 2 studies)	MRM/SRM/PRM assays (NCI assay portal)
		Log2[FC] Class1 / Class3	Log2[FC] MIBC / NMIBC	2vs3 Log2(FC)*(-1)	GU vs Urobasal(A) LUND		
S100A9	Protein S100-A9	4.40	1.94	1.40	1.16	yes	yes
PDIA4	Protein disulfide-isomerase A4	0.76	0.58	0.84	0.41	yes	yes
TFRC	Transferrin receptor protein 1	1.75	0.78	0.94	0.65	yes	yes
AARS	Alanine-tRNA ligase, cytoplasmic	1.92	1.04	0.54	0.41	yes	yes
G6PD	Glucose-6-phosphate 1-dehydrogenase	1.11	0.80	0.52	0.27	yes	yes
HSP90B1	Endoplasmic	0.75	0.51	0.64	0.25	yes	yes
DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	0.33	0.97	0.91	0.12	yes	yes
CTSZ	Cathepsin Z	1.73	1.35	0.97	0.60	yes	yes
RPN1	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	0.53	0.44	1.02	0.44	yes	yes
GRB2	Growth factor receptor-bound protein 2	3.09	1.90	0.84	0.28	yes	yes - 1 out of the 6 assays used enrichment (peptide immunoaffinity in plasma)
HK1	Hexokinase-1	-0.53	-0.37	-0.43	-0.57	yes	yes
ANXA1	Annexin A11	-0.64	-1.04	-0.73	-0.41	yes	yes
CAPN1	Calpain-1 catalytic subunit	-0.78	-0.88	-0.47	-0.25	yes	yes
USP5	Ubiquitin carboxyl-terminal hydrolase 5	-0.39	-0.32	-0.42	-0.13	yes	yes
BCAM	Basal cell adhesion molecule	-1.25	-1.89	-0.88	-0.56	yes	yes

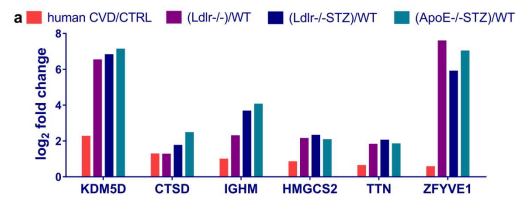


Proteomic analyses of thoracic aortas from animal models

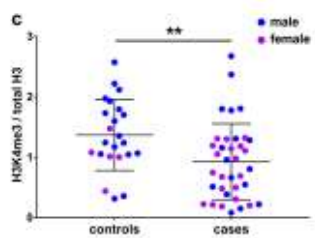


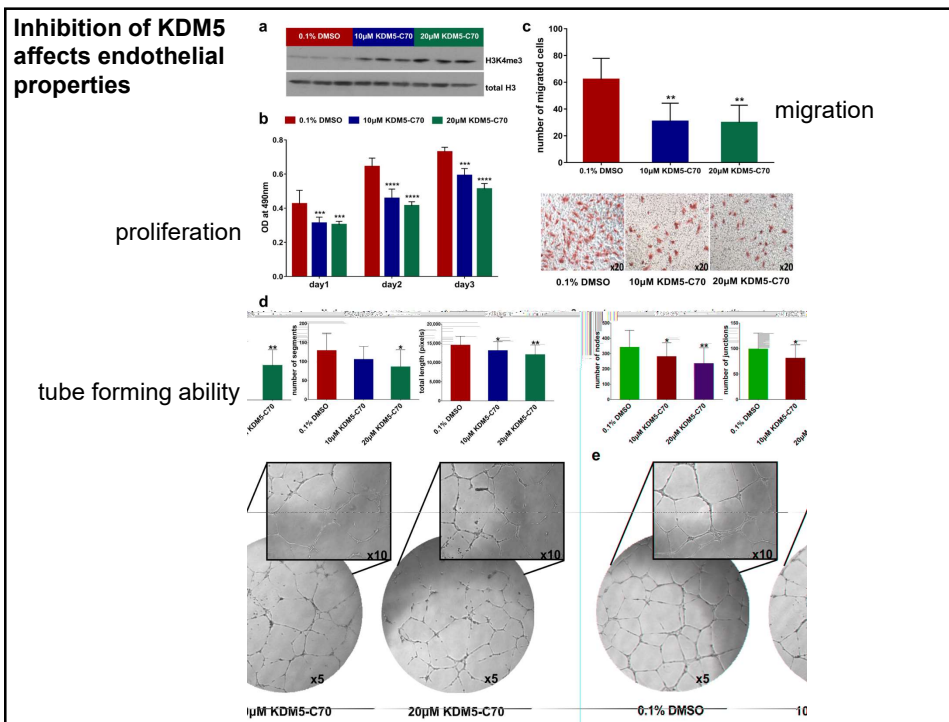
Cross species comparisons shortlist conserved CVD-associated changes

Conserved pronounced change of KDM5D in human and animal CVD



Concomitant change of KDM5D substrate (H3K4me) in human CVD





In summary

- **Biomarker development** is associated with multiple challenges
 - Discovery (clinical and statistical design, data validation and interpretation)
 - Validation (assay optimization, clinical resources)
 - Implementation (multi-disciplinary interactions, establishment of cost effectiveness, added-value..)
- Current efforts focus on **integrating data from multiple sources** (clinical, various -omics, literature) for the identification of "biology-driven" markers and targets (personalized medicine).
- **Marker panels** are better reflective of disease heterogeneity and tolerate instability and inconsistency of individual biomarkers
- Multi-omics analysis increases power and validity of individual observations

Integrate –omics to develop personalized treatment approaches

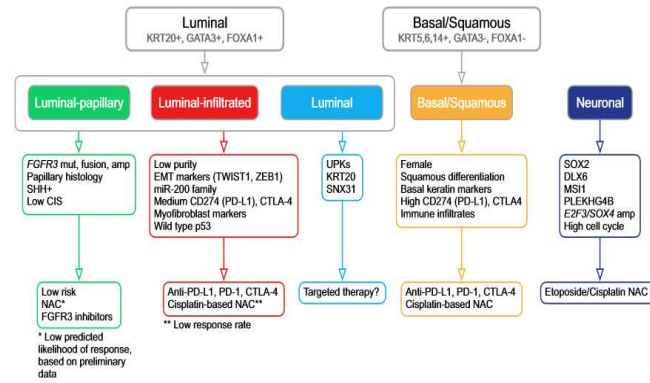


Figure 7. Proposed Schema of Expression-Based, Subtype-Stratified Therapeutic Approach as a Framework for Prospective Hypothesis Testing in Clinical Trials

Robertson et al., Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer, Cell (2017), <https://doi.org/10.1016/j.cell.2017.09.007>

Functional -omics in a systems biology approach



**Data Mining
Backfilling**



**Exp. Data
Collection**



Simulation

Creating the full story from snapshots is not easy....



ACKNOWLEDGEMENTS

BRFAA

Manousos Makridakis
Marika Mokou
Rafael Stroggilos
Jerome Zoidakis
Georgia Kontostathi
Vassiliki Lygirou
Elena Critselis
Agnieszka Latosinska (DE)
Maria Frantzi (DE)

Laboratory of Gene Therapy

Maria Roubelakis
Nikolaos Anagnou

ATHENS MEDICAL SCHOOL

K. Stravodimos
A. Bamias
A. Lazaris
K. Doulmas

European Collaborators

Harald Mischak (MOS, DE)
William Mullen (GLA, UK)
Joost Schanstra (INSERM FR)
Julie Klein (INSERM,FR)
Nuria Malats (CNIO, ES)
Ellen Zwarthoff (ERASMUS, NL)
Walter Kolch (UCD, IR)
Joachim Jankowski (UKA, DE)
Griet Glorieux (Ghent U, BE)
Ray Vanholder (Ghent U, BE)
Vincent Jacquet (U Geneva, CH)



Thank you

vlahoua@bioacademy.gr



