# Bridging the gap: From hypothesis-independent tests to understanding of biological mechanisms

H Roder, CTO, Biodesix

# The Importance of Protein Expression

## Complex diseases raise intrinsically phenotypical issues

***proteomics vs. genomics*** ***phenotype vs. genotype***

We need to measure the actual state of the organism when disease occurs.

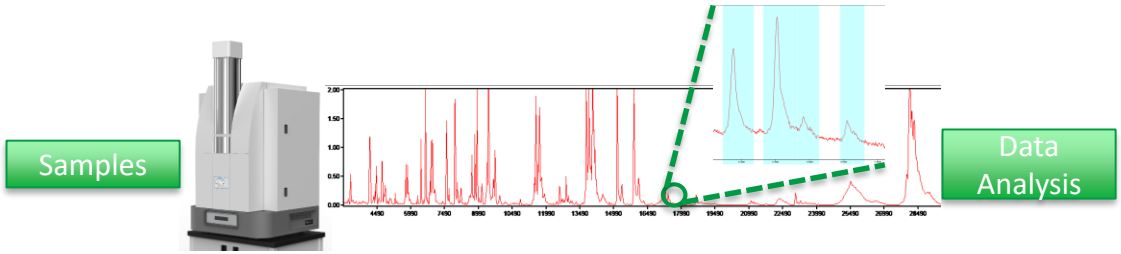
In order to build tests based on circulating markers we should look at protein expression in serum/plasma

biodesix®

# The measurement

## From MALDI-MS profiling

- Clinically relevant tests (not biomarker research)
  - Design goals:
    - High throughput: Fast turnaround times ( >100 samples/day)
    - High reproducibility: Both on a feature (CV <6%) and label level (>90% conc)
    - Ease of use: Don't change clinical practice (serum), easy logistics (not frozen)




- The data for test development
  - N = 50-1000 samples
  - P = 300-600 attributes
  - Clinico-pathological data
  - Outcome (time to event, ORR)
  - **Biology agnostic, hypothesis independent**

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | Ftr1 | Ftr2 | Ftr3 | Ftr4 | Ftr5 | Ftr6 | Ftr7 | Ftr8 | Ftr100 | TrainingLabel | |
| 2 | N1 | 0.415971 | 0.453317 | 0.673903 | 0.076092 | 0.147629 | 0.12599 | 0.992154 | 0.321276 | 0.446342 | 0 | |
| 3 | N2 | 0.914014 | 0.828998 | 0.419234 | 0.835347 | 0.028113 | 0.141281 | 0.226954 | 0.22908 | 0.038007 | 1 | |
| 4 | N3 | 0.598595 | 0.194144 | 0.625446 | 0.875984 | 0.872268 | 0.547563 | 0.174016 | 0.475074 | 0.133319 | 0 | |
| 5 | . | 0.847929 | 0.865962 | 0.824697 | 0.568986 | 0.940735 | 0.747775 | 0.887919 | 0.832525 | 0.938594 | 0 | |
| 6 | . | 0.776031 | 0.43532 | 0.441537 | 0.085391 | 0.687222 | 0.813322 | 0.276762 | 0.695122 | 0.095569 | 1 | |
| 7 | . | 0.570988 | 0.137728 | 0.753186 | 0.010329 | 0.688717 | 0.484652 | 0.048171 | 0.056074 | 0.866319 | 1 | |
| 8 | . | 0.641197 | 0.837823 | 0.266138 | 0.492866 | 0.330169 | 0.057167 | 0.550032 | 0.851663 | 0.610077 | 1 | |
| 9 | Nn | 0.799969 | 0.25303 | 0.521209 | 0.676472 | 0.957106 | 0.742081 | 0.627158 | 0.44552 | 0.131462 | 0 | |

# Established: VeriStrat

## MALDI-TOF MS can be used in clinical practice

VeriStrat assigns VeriStrat labels, Good and Poor, to a patient's serum sample.

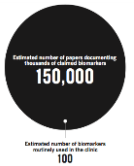
Veristrat is a truly multivariate proteomic marker used in NSCLC

- Veristrat is prognostic in multiple lines of therapy, and predictive for targeted therapy vs chemotherapy (validated in a prospective PIII study, PROSE)
- Veristrat is orthogonal to genomic measurement and provides independent additional information on host response to a tumor
- VeriStrat Poor patients have elevated levels of acute response proteins that are understood to be markers of poor prognosis
  - The elevation of which can lead to activation of MAP/K leading to "resistance" to some targeted therapies targeting upstream receptors.
- No cut-off of a single univariate VS component works as well as VeriStrat
  - Smaller effect sizes
  - Poorer reproducibility

Since Veristrat, we have made substantial progress

- By measuring the abundance of many more proteins than previously possible
- By taking advantage of paradigm shifting advances in modern machine learning

# The p>N problem: “deep data”




Poste, 2011

## Many attributes + multivariate = many, many classifiers

- With many attributes which we can combine in many ways we can build very many classifiers
  - By random chance some of these will show good performance on the data set
  - Hopefully we have more potentially useful classifiers than expected by random luck alone (but even if not, there could still be something useful there).
  - Feature selection easily overfits
- Which one to take?
  - Look to see which ones show potential for adequate performance (there will probably still be a lot) and use all of them

**Boosting: combining a set of weak learners into a learner with at least as good or better performance**

**Tuning/classifier design: picking our definition of “adequate performance” to suit clinical need**

biodesix

# Classifier development

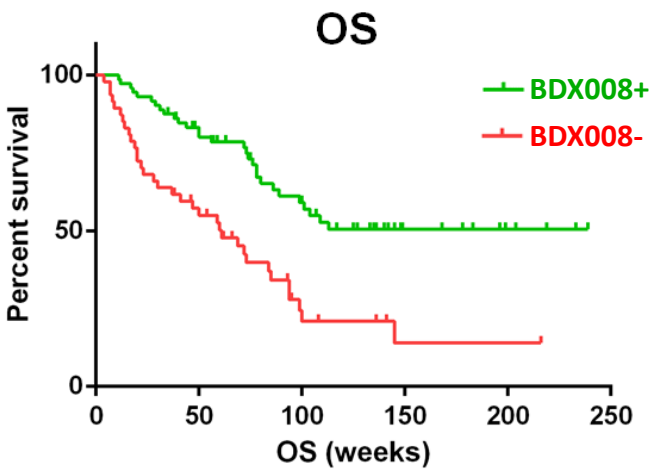
## dxCortex™

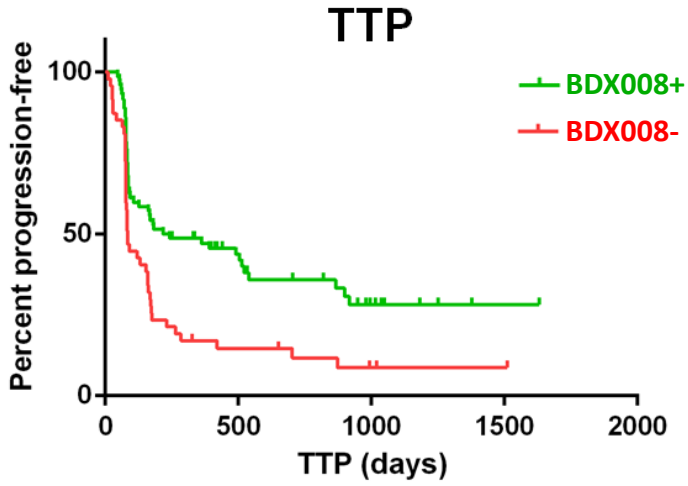
Biodesix classifier development uses a multivariate machine learning approach borrowing ideas from “deep learning” to correlate the measured molecular data (protein abundances) with clinical outcome.

- The dxCortex was designed to deal with the problem of having many more attributes p (protein abundance) than instances N(samples), the p>>N problem, without overfitting.
  - In many studies we have shown that development set performance estimates were validated in independent blinded sets.
- The dxCortex can be guided to solve defined clinical problems
  - For example to build a predictive test
- The dxCortex generates clinical group labels
  - Using iterative techniques to produce robust stratification
- The dxCortex does not require the selection of particular features
  - It works well without needing to introduce bias.

# In development: Pre-tx serum mass spectra predict benefit from anti-PD1: BDX008

- Pre-treatment serum samples from melanoma patients treated with nivolumab
- The test assigned 72 patients (61%) a BDX008+ classification and 47 (39%) a BDX008- classification and validated on independent blinded sets
- Patients classified as BDX008+ had significantly better overall survival (OS) and time-to-progression (TTP) than those classified as BDX008-



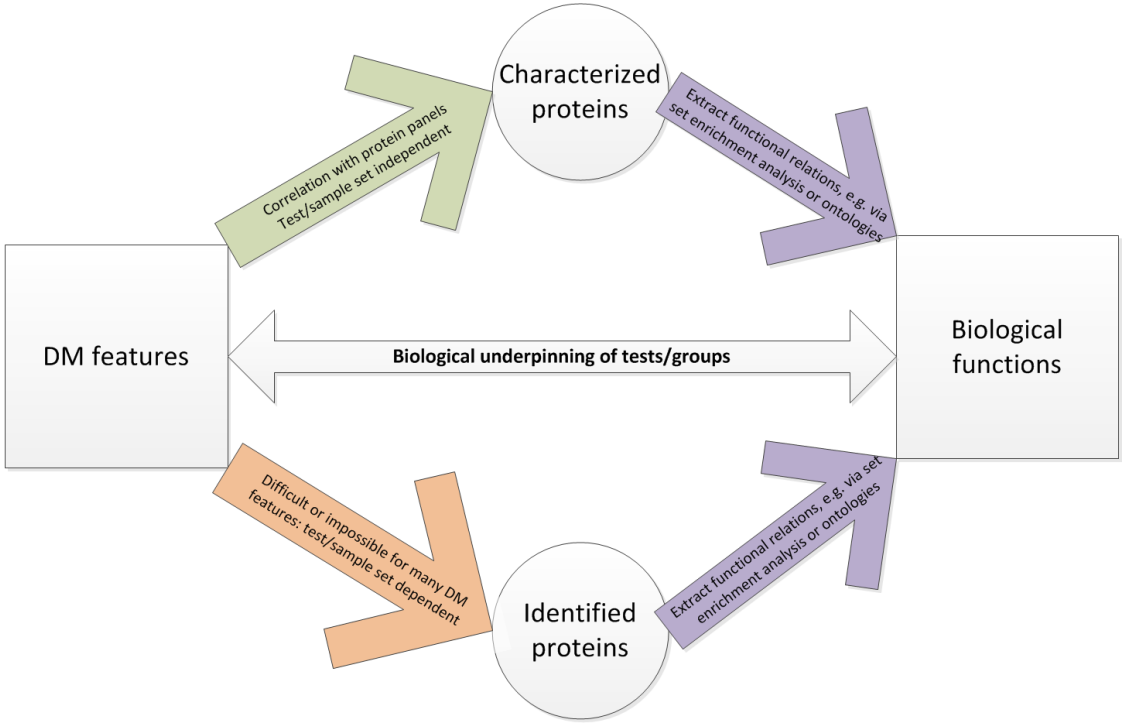

| | OS | TTP |
|---|---|---|
| HR (95% CI) | 0.38 (0.19-0.55), p<0.001 | 0.50 (0.29-0.71), p=0.001 |
| Median BDX008- | 61 weeks | 84 days |
| Median BDX008+ | Not reached | 230 days |

J Weber et al., *Pre-treatment patient selection for nivolumab benefit based on serum mass spectra*. SITC2015.

biodesix

# MALDI features and biological functions

## Can we understand multivariate tests from a biological point of view?

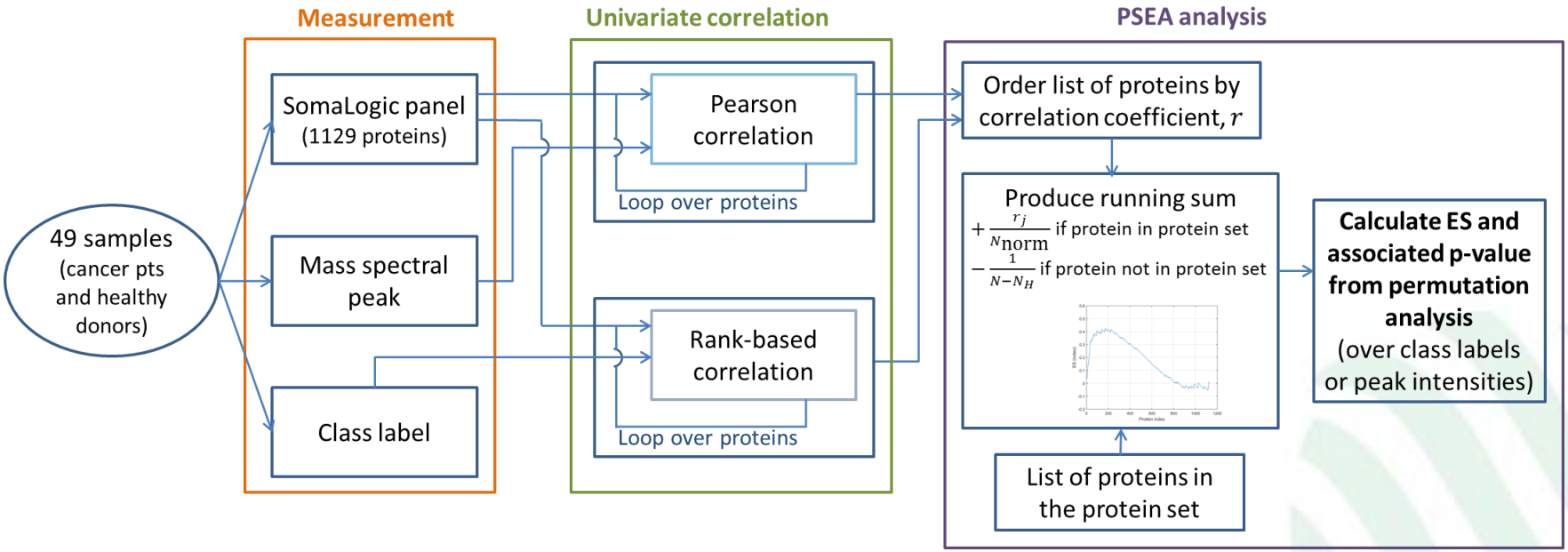



- Protein ID's are hard to impossible
  - Even if you have IDs, you still need to relate to functions.
- Can we use correlational techniques to avoid using protein IDs?

biodesix

# Use GSEA analysis for protein expression

- One-time experiment to measure the expression of well-characterized proteins
- Define biological processes of interest from these measured proteins
- Generate:
  - Association of these processes with test labels (rank correlation)
  - Association of these processes with individual features (Pearson)



# Correlation of Protein Sets with BDX008 labels

| Protein sets | Enrichment score | p value |
|---|---|---|
| **Acute inflammatory response** | 0.424 | 0.021 |
| Activation of innate immune response | 0.412 | 0.551 |
| Regulation of adaptive immune response | -0.234 | 0.903 |
| Positive regulation of glycolytic process | -0.495 | 0.295 |
| Immune T-cells | -0.156 | 0.975 |
| Immune B-cells | 0.213 | 0.907 |
| Cell cycle regulation | -0.207 | 0.813 |
| Natural killer regulation | -0.406 | 0.392 |
| **Complement system** | 0.552 | 0.011 |
| Acute response | 0.539 | 0.098 |
| Cytokine activity | -0.231 | 0.666 |
| **Wound healing** | -0.373 | **0.099** |
| Interferon | -0.178 | 0.942 |
| Interleukin-10 | 0.190 | 0.763 |
| Growth factor receptor signaling | -0.221 | 0.453 |
| **Acute phase** | 0.572 | 0.011 |
| Hypoxia | -0.247 | 0.650 |
| Cancer | 0.153 | 0.958 |

- At the $p < 0.05$ level there are correlations of the class labels with the protein sets corresponding to acute inflammatory response, acute phase, and complement system.

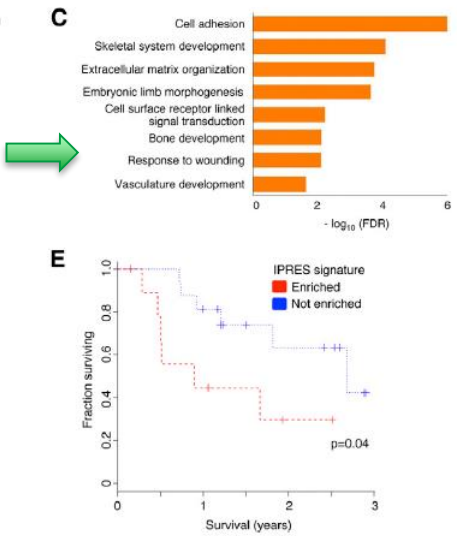
- At the $p < 0.10$ level there are also correlations of the class labels with wound healing.

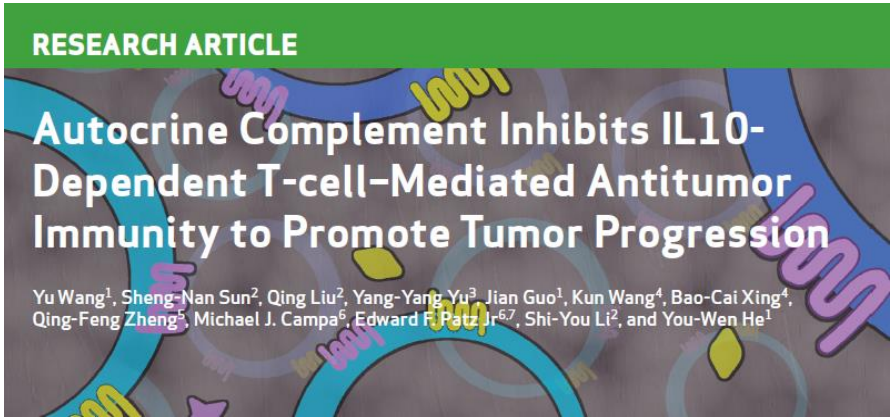
biodesix

# Independent Validation

## Recent papers identify complement and wound healing pathways as key in modulating response to checkpoint inhibitors

Combined blockade of complement signaling and anti-PD-1 can enhance anti-PD-1 efficacy; Cancer Discovery 6 (9) :1022-35 June 2016



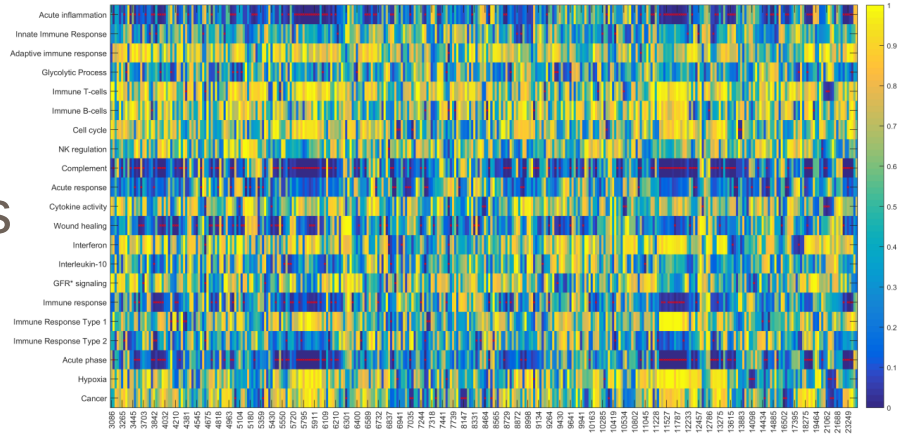




A transcriptional signature (IPRES) identified related to innate anti-PD-1 resistance; Wound Healing is one of the pathways; Cell 165: 35-44 March 2016

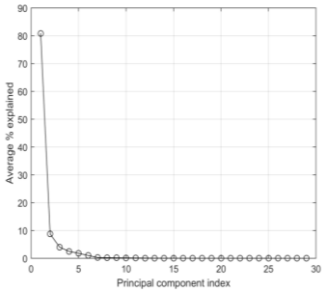
***Biodesix independently identified these key pathways***

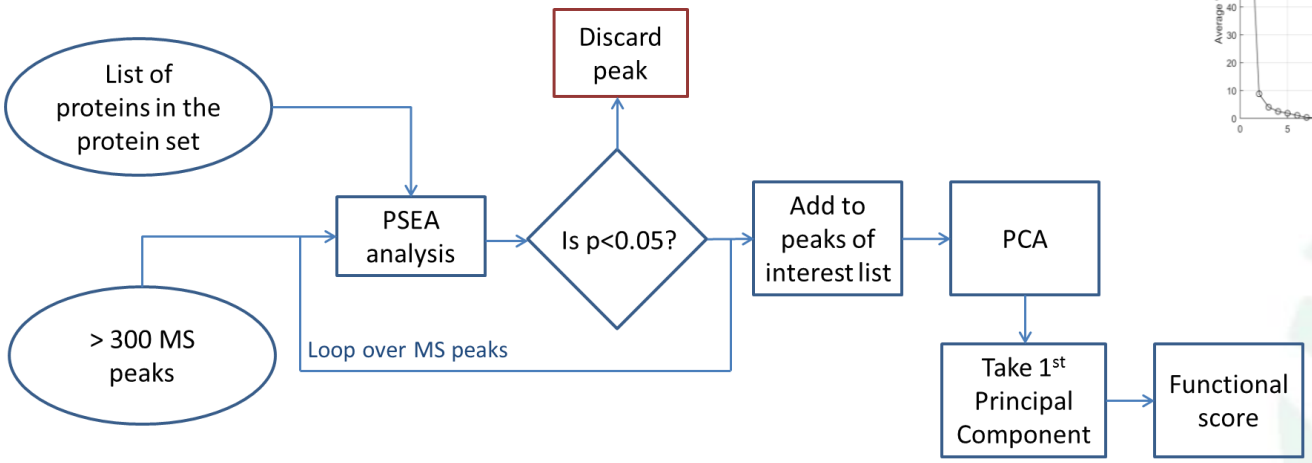
# Phenotypical scores related to processes

## From mass spectrometry data




- Peaks → biological process
- PCA on a population
  - Reference set, e.g. NSCLC
  - Use first principal component for each process to describe
  - Validate



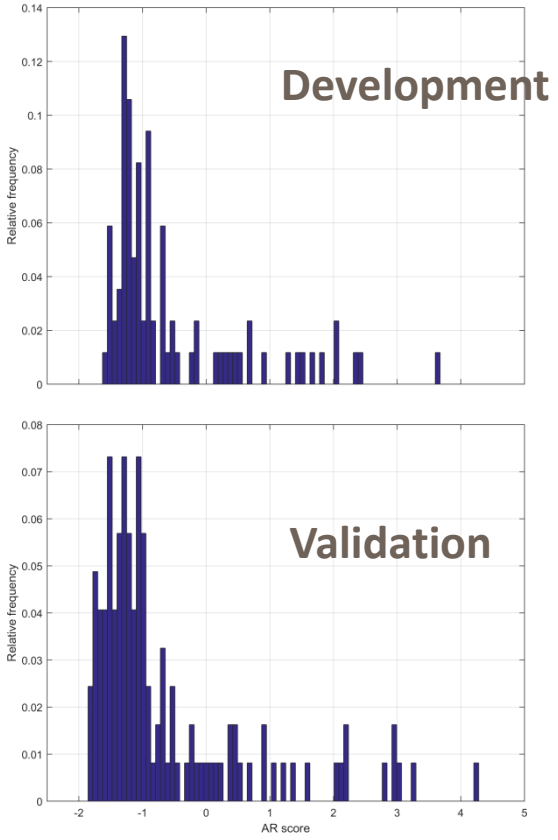




biodesix

# PSEA score validation

## Develop on one set, compare to another

- Sets
  - From randomized PHIII study (PROSE)
  - Development from arm 1 (N=85)
  - Validation from arm 2 (N=123)
- For example for acute response (AR)
  - Relevant for immunotherapies
- The score distribution validates
- Other scores
  - Wound healing (WH), complement (C)

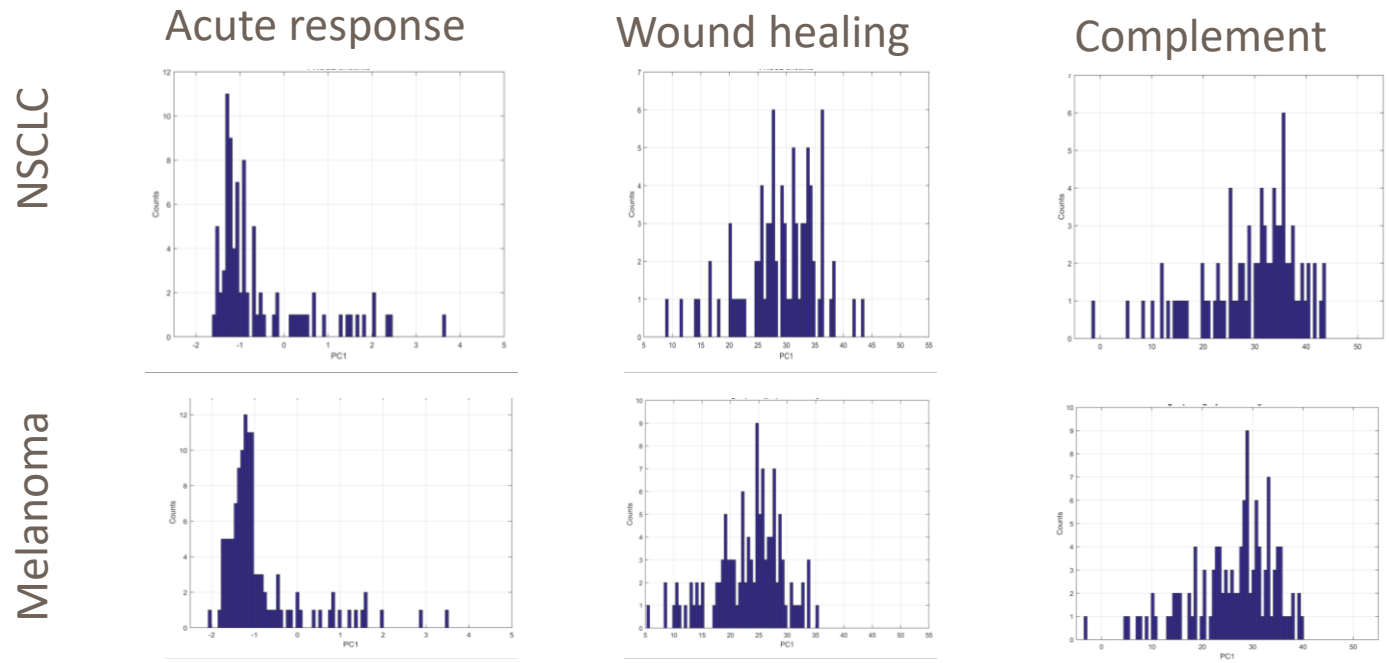



| Percentile | Development | Validation |
|---|---|---|
| 10 | -1.40 | -1.65 |
| 20 | -1.29 | -1.50 |
| 25 | -1.25 | -1.46 |
| 30 | -1.23 | -1.34 |
| 40 | -1.14 | -1.25 |
| 50 | -0.98 | -1.10 |
| 60 | -0.89 | -0.98 |
| 70 | -0.59 | -0.68 |
| 75 | -0.20 | -0.33 |
| 80 | 0.27 | 0.08 |
| 90 | 1.41 | 1.24 |

biodesix

# What can we do with PSEA scores?

## Compare indications: NSCLC and metastatic melanoma



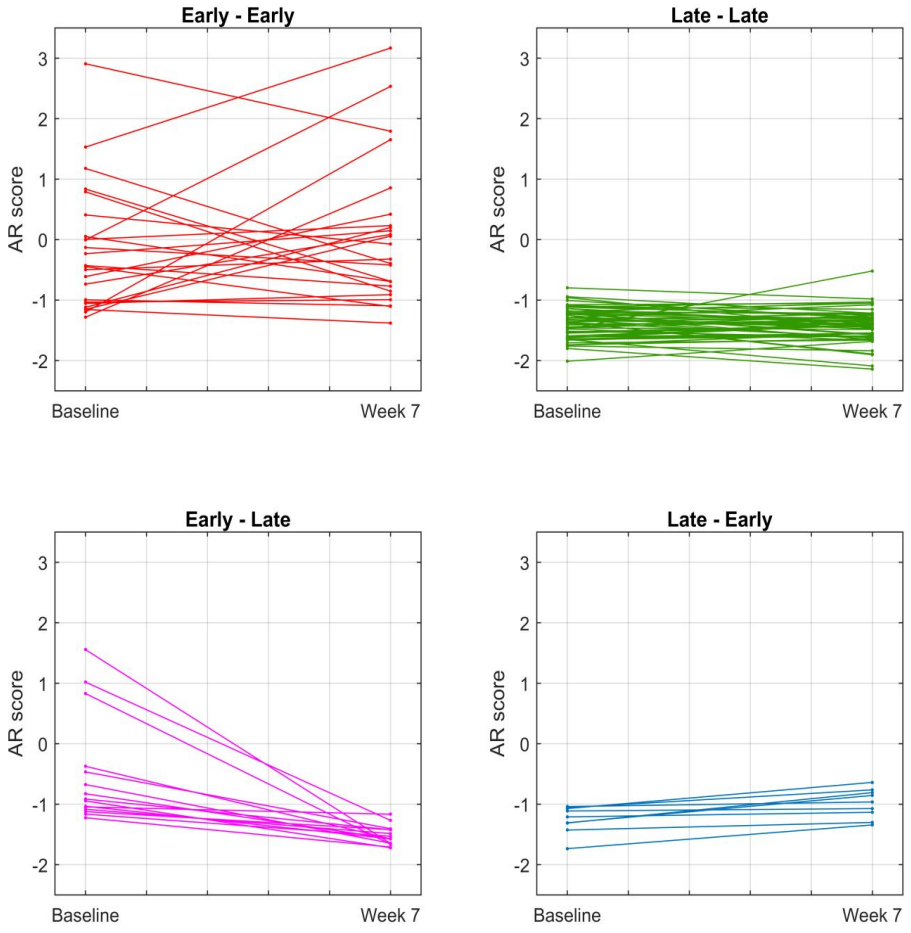

- AR is very similar between NSCLC and melanoma.
- WH and C are different
  - Immunosuppressive TME; anti-CTLA4/anti-PD1 combos more effective in NSCLC; possible benefit of anti MDSC tx ?

biodesix

# Concept: Process monitoring using scores

## AR score for BDX008 longitudinally

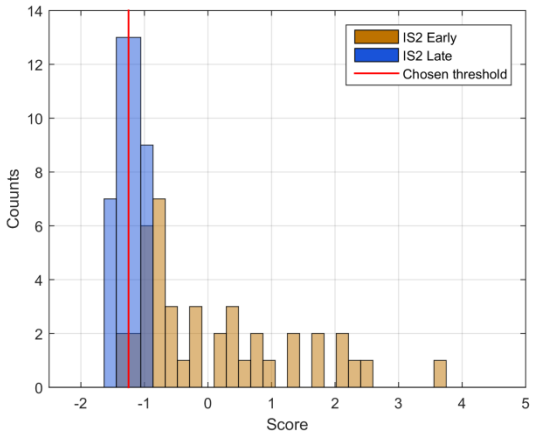
- Changes from Early to Late are accompanied by a decrease in AR score
- Changes from Late to Early are accompanied by an increase in AR score.
- Those that remain Early show no systematic behavior.
- Those that remain Late show little change in AR score

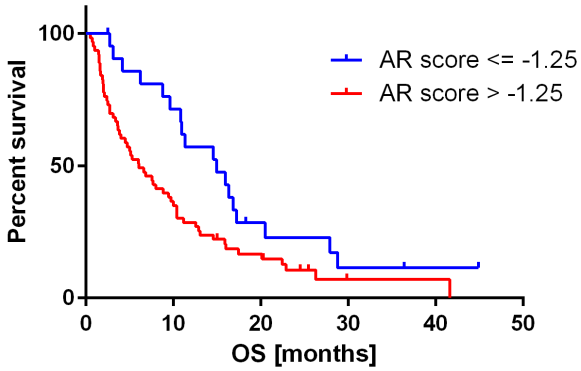


# Closing the circle

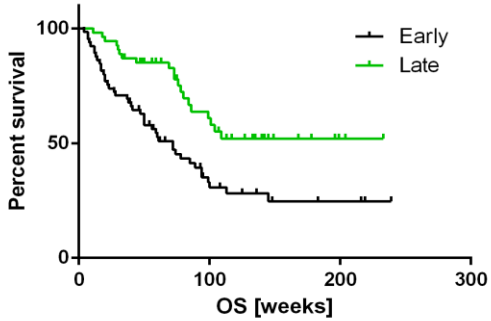
## Tests using scores are possible

- From an AR cut-off based




- Using AR, WH, and C




- It works, and provides a biological foundation
  - But agnostic tests still have advantages

# Discussion

Based on simple blood draws

- One can develop clinically useful multivariate tests using correlative tools adapted from the artificial intelligence world
- These can be interpreted in terms of (coarse) biological functions
  - which give insight into the relative importance of these functions across indications
  - Which has been confirmed and rediscovered using mRNA measurements
    - For acute response, wound healing, complement
- This approach also enables the phenotypical profiling and monitoring of patients
- The measurement of processes other than the three presented is possible using biological insight
  - For example the IL10 cascade and glycolytic processes
- Tests based on biological scores are possible
  - But still require the measurement of all the proteins contained
  - And are not as robust as agnostically developed tests
- Many improvements are possible
  - Larger and different reference sets
  - Methodological improvements, e.g. KPCA
  - **Collaborations are welcome**

# Acknowledgements

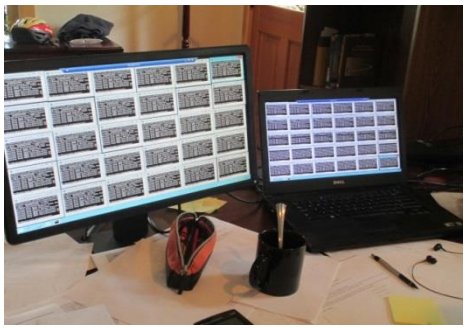
This would not have been possible
without the support of

Patients and their families

Our collaborators

The Biodesix R&D team

Contact: heinrich.roder@biodesix.com



Our computers