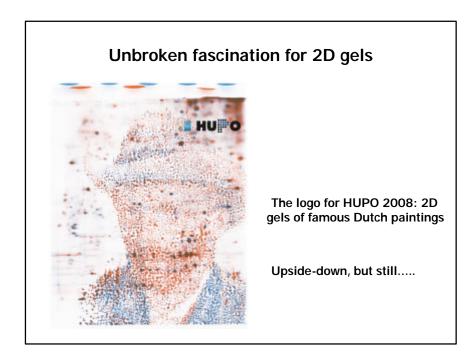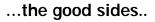### *Reproducibility of 2D gel-based proteomics experiments*
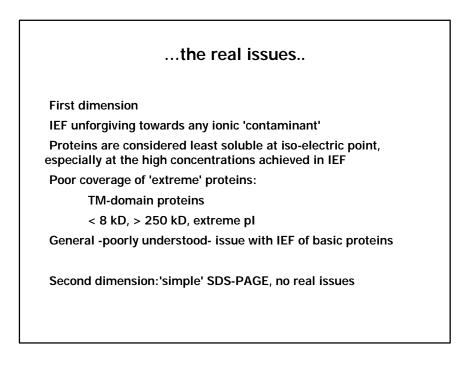
David Bramwell[1], Mary Caponite Hurley[2], Alamgir Khan[3], Katrin Marcus[4], Jules A. Westbrook[5] ,Hans Voshol[6]
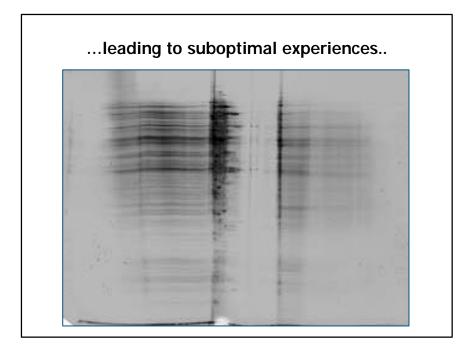
[1] Nonlinear Dynamics
[2] Michigan Proteome Consortium.
[3] Australian Proteome Analysis Facility Ltd
[4] Medical Proteom-Center Bochum
[4] University College Dublin
[6] Novartis Institutes for BioMedical Research

Unbroken fascination for 2D gels

The logo for HUPO 2008: 2D gels of famous Dutch paintings

Upside-down, but still…..
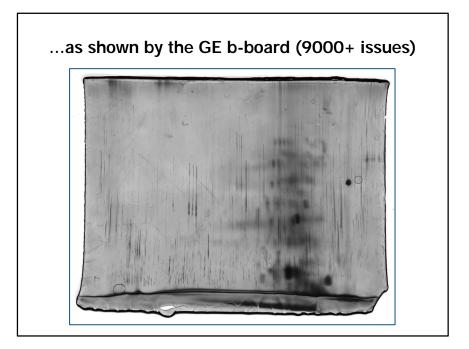
...the good sides..

- Limited sample prep required and denaturing conditions ® sample preservation
- Inherently reproducible subset of proteome: always the 2000 most abundant protein species in the sample (in special gel systems even more)
- Instant quantitation with good dynamic range using fluorescent dyes
- Straightforward ID and (partial) characterization of proteins, since all peptides in 1 spot
- Excellent resolution of most isoforms
- Parallel sample processing

...the real issues..

First dimension

IEF unforgiving towards any ionic 'contaminant'

Proteins are considered least soluble at iso-electric point, especially at the high concentrations achieved in IEF

Poor coverage of 'extreme' proteins:

    TM-domain proteins

    < 8 kD, > 250 kD, extreme pI

General -poorly understood- issue with IEF of basic proteins

Second dimension:'simple' SDS-PAGE, no real issues

...leading to suboptimal experiences..

...as shown by the GE b-board (9000+ issues)

Challenges in comparing differential proteomics experiments

- **Coverage of the proteome**
  - **if proteome coverage is partial in each experiment, overlap between them is inherently limited**
- **Technical challenges, e.g. sampling in LC-MS experiments**
  - **'inherent' lack of reproducibility: dependent on sample complexity**

- **2D-PAGE**
  - **no sampling problem – inherent reproducibility**
  - **many technical challenges**

- **Goal of this experiment**
  - **short-term: assess sources of variability**
  - **long-term: provide reference materials and protocols to proteomics community**

Validating 2D PAGE in practice
sources of variability

sample preparation   *identical samples*

↓

IEF

↓   *strips from 1 batch*

SDS-PAGE   *detailed protocols*

↓

staining

↓

scanning

↓

image analysis with more or less user intervention

↓   *less stringent instructions*
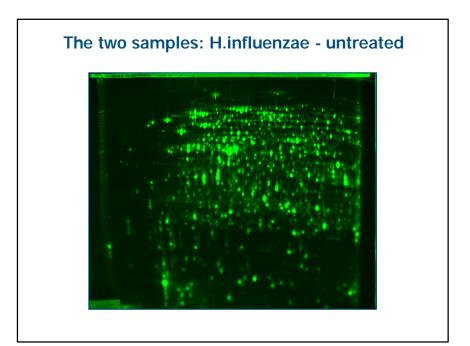
statistical analysis

↓

reporting results

Protocol reminder

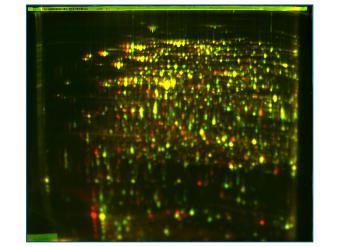## In the original invite the protocols were stated as:

"Each participating lab should run the two different, but related, samples on 2-D gels according to the supplied protocol, then perform an analysis between the two groups, using the software provided to identify what are in their opinion the top 200 significantly changing spots. (This would be up to 200 if the lab believes that there are less than 200 significantly changing spots, or over 200 spots if this result is obtained.) This of course includes newly appearing spots.

In principle no spot editing should be required, but if deemed needed any rejected spots should have comments added as to why they were rejected or edited. (This is easily done in the supplied workflow.)
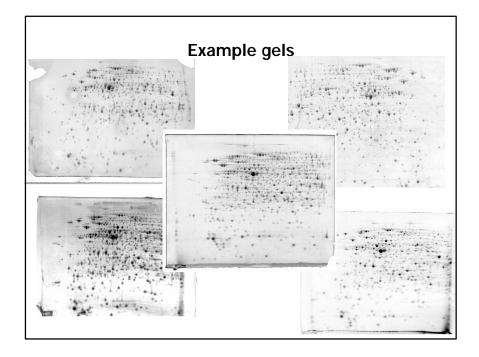
The full analysis should then be archived and uploaded to the specified site for comparison between labs and further analysis."
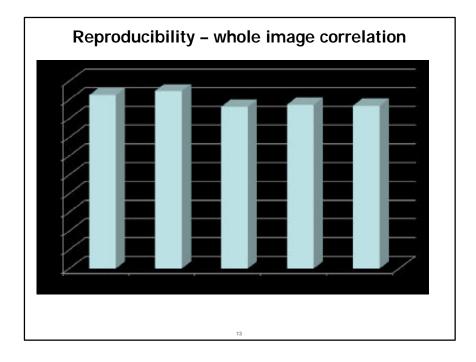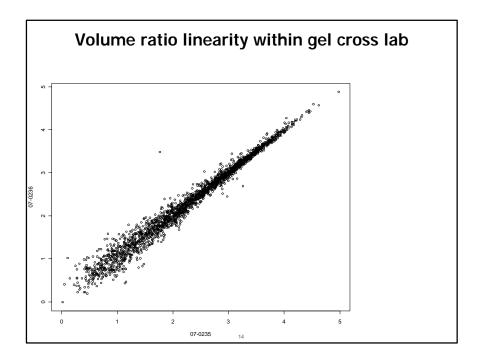

The two samples: H.influenzae - untreated

and treated with actinonin (a peptide deformylase inhibitor

hundreds of differences to challenge the matching/alignment process

Example gels

Reproducibility – whole image correlation

Volume ratio linearity within gel cross lab

07-0235

Volume ratio – cross lab

07-0235 15

Reported top 50 differential spots

	unconstrained user analysis	NLD automatic analysis
user1	92	98
user2a	84	98
user2b	76	100
user2c	88	100
user3	70	88
user4	88	84
user5a	90	94
user5b	94	82
user5c	94	94

As reference/gold standard a constrained analysis of each set by a professional analyst was used

**Variability of multiple runs
(1 lab, 3 operators, 4 week timeframe)**

20 gels each of treated and non-treated samples

High reproducibility does not require internal standards

Conclusions

- **Protein expression analysis based on 2D PAGE is reproducible across labs (and highly reproducible within a lab)**
  - labs could easily share a common reference and hence the linked ID's

- **Main sources of variability**
  - user manipulation and interpretation of images (this study)
  - sample prep (in general)??

- **With some constraints, experimental procedures appear to be remarkably robust**
  - most similar images were generated with three different IEF instrument
  - variability in 2nd dimension, which affects direct image alignment, but not differential analysis

Next steps

- **Expand the experiment beyond the original five labs (ongoing)**

- **Submit study results for publication (finally....)**

- **Provide reference images and protocols to community – done, see fixingproteomics.org**

- **Provide reference sample to community**

- **Further options under discussion**
  - central image repository with matching function for web-based QC of gels with reference sample

Acknowledgements

- **Teams in Bochum, Sydney, Dublin, Ann Arbor, Basel for producing the data**

- **Team at Nonlinear in Newcastle for a large effort in coordinating and performing the analysis**

- **Bio-Rad and CILBiotech for supporting the next phase with samples and logistics**

- **Jan van Oostrum and Will Dracup, the motivators behind (and sometimes at the forefront of) the scenes**