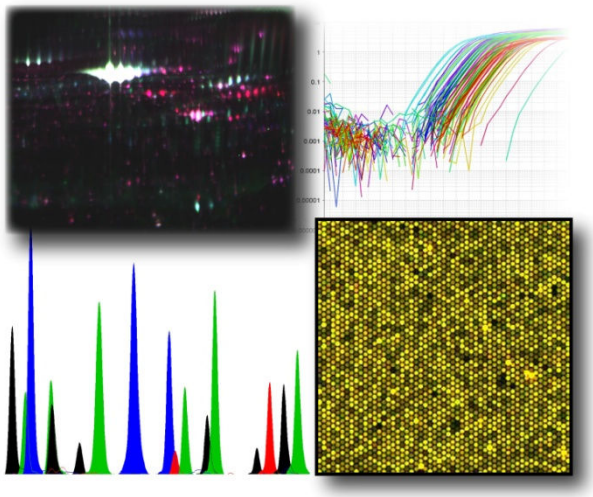
MACPROGEN

2nd Genomics and Proteomics Workshop

June 27-29, 2011

Skopje, Republic of Macedonia



MACPROGEN

2nd Genomics and Proteomics Workshop

June 27-29, 2011

Skopje, Republic of Macedonia

Organized by:

Research Center for Genetic Engineering and Biotechnology, Macedonian Academy of Sciences and Arts

Sponsored by:

FP7 grant "National Reference Centre for Genomics and Proteomics" (acronym – MACPROGEN), funded by the Seventh Framework Programme, REGPOT-2008-1.

SEVENTH FRAMEWORK PROGRAMME

CONTENTS MANUEL

PROGRAM

June 27, 2011	***Monday***
Venue	**RCGEB**
Participants	**RCGEB scientists and invited speakers**
12.00 – 16.00	Practical session (held in parallel):
	Microarray technology:
	Designing and performing microarray experiments
	Troubleshooting hybridisation and washing
	2D DIGE technology: *DIGE experimental design*
18.00	*Dinner*

June 28, 2011	***Tuesday***
Venue	**MASA meeting rooms**
Participants	**RCGEB scientists and invited speakers**
09.00 – 09.15	Dijana Plaseska-Karanfilska: *RCGEB and MACPROGEN project overview*
09.15 – 10.45	Short talks and discussion:
	Katarina Davalieva: *DIGE analysis of breast cancer tissues*
	Lyubomira Chakalova : *Non-coding RNA profiling by RNA-Seq*
	Svetlana Madjunkova: *Early and noninvasive molecular prenatal diagnosis*
10.45 – 11.00	*Coffee break*
11.00 – 12.00	Practical sessions (held in parallel):

Microarray technology: *Data analysis and web resources*

Real-time PCR: *CNV validation by real-time PCR*

2D DIGE technology: *Practical aspects of performing a DIGE experiment*

12.00 – 13.00 *Lunch break*

Venue **MASA amphitheatre**

Participants **Representatives of the Macedonian biomedical community**

13.00 – 13.15 Dijana Plaseska-Karanfilska: *MACPROGEN overview*

13.15 – 14.00 Thilo Dork, Germany: *Molecular genetics of breast cancer – an update*

14.00 – 14.45 Neil Avent, UK: *Using DIGE to define new markers for non-invasive prenatal diagnosis*

14.45 – 15.30 Paul Brady, Belgium: *Genomic arrays in diagnostics*

June 29, 2011 ***Wednesday***

Venue **RCGEB**

Participants **RCGEB scientists and invited speakers**

09.00 – 12.00 Practical sessions (held in parallel):

Microarray technology: *Custom array design*

Real-time PCR: *Gene expression analysis*

2D DIGE technology: *DIGE data analysis*

Closing

SPEAKER LIST

Thilo Dörk-Bousset

Hannover Medical School,

Gynaecology Research Unit,

Hannover, Germany

Neil Avent

School of Biomedical and Biological Sciences,

Faculty of Science and Technology, University of Plymouth,

Plymouth, Devon, UK

Paul Brady

Labaratory for Cytogenetics and Genome Research,

Department of Human Genetics, Biomedical Sciences Group,

K.U.Leuven, Leuven, Belgium

PARTICIPANTS

Dijana Plaseska-Karanfilska

Emilija Sukarova-Stefanovska

Katarina Davalieva

Lybomira Chakalova

Predrag Novevski

Sanja Kiprijanovska

Ivana Maleva

Svetlana Madzunkova

Mirko Trpevski

Zvezdana Moneva

Jugoslav Stojcevski

Gorgi Bozinovski

TALKS

Order:

1. Dijana Plaseska-Karanfilska: *RCGEB and MACPROGEN project overview*
2. Katarina Davalieva: *DIGE analysis of breast cancer tissues*
3. Lyubomira Chakalova : *Non-coding RNA profiling by RNA-Seq*
4. Svetlana Madjunkova: *Early and noninvasive molecular prenatal diagnosis*
5. Thilo Dork, Germany: *Molecular genetics of breast cancer – an update*
6. Neil Avent, UK: *Using DIGE to define new markers for non-invasive prenatal diagnosis*
7. Paul Brady, Belgium: *Genomic arrays in diagnostics*

"National Reference Centre for Genomics and Proteomics"

2nd Genomics and Proteomics Workshop

[redacted]

MACPROGEN PROJECT

FP7-REGPOT-2008-1 Grant Agreement No. 229458

RCGEB Background

- Founded in 1986 as a research unit of MASA with the main goal of advancing scientific knowledge in the field of protein chemistry, molecular biology, genetic engineering and biotechnology through research, practical training of scientists, postgraduate studies and application
- The National Reference Laboratory for Hemoglobinopathies founded in 1970 has become a part of RCGEB
- RCGEB "Georgi D. Efremov"

RCGEB

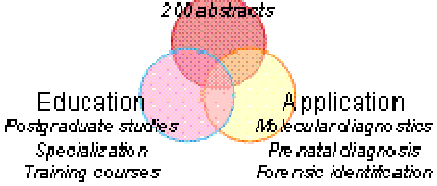
- the ***first*** institution in ***Former Yugoslavia*** that introduced ***DNA technologies*** and was for many years at the ***forefront*** of the research, application and education of these disciplines in the region
- the ***major research, education and diagnostic facility*** in the field of ***molecular genetics*** in R. Macedonia which has reached national, regional and international recognition
- one of the few institutions from Macedonia that met the criteria for ***Centre of excellence*** due to its scientific and educational record in life sciences.

RCGEB

- ***Space***
  500 m² laboratory
  200 m² offices & library
- ***Staff***
  Scientific staff
  Administrative / Technical staff
- ***Equipment***
  Microarray facility, Automated sequencers, Real-time PCR, Thermal cyclers, Luminometer, Thermal shaker, 2D-DIGE, Amino acid analyzer, HPLC, Ultracentrifuge, Spectrophotometers, Scintillation counter, Water purification system, GelDoc system

RCGEB Activities

Research
30 projects
150 papers
200 abstracts



Research projects

Molecular characterization of monogenic diseases
- *Hemoglobinopathies; Cystic fibrosis; Hemophilia; Muscular dystrophy; Spinal Muscular Atrophy; Fragile X syndrome; Huntington's disease; Cystinuria;*

Molecular epidemiology of infectious diseases
- *HPV, HCV, HBV, Chlamydia trachomatis*

Molecular basis of malignant diseases
- *Colon; Breast & ovarian; Prostate; Lung cancer*

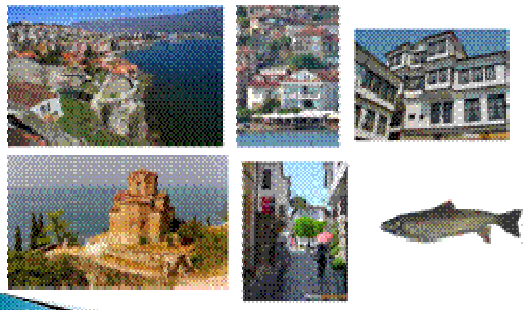
DNA markers *(human identification & population studies)*

Reproductive genetics
- *Male infertility, missarriages, prenatal diagnosis*

RCGEB – Research Funds

- Macedonian –American Joint Funds for Scientific & Technological cooperation
- National Institute of Health of the USA
- American Ministry of Agriculture
- ICGEB – Trieste, Italy
- European Commission
- Ministry of Science of the R. Macedonia
- Science funds of the Macedonian Academy of Sciences and Arts

OHRID



Balkan Journal of Medical Genetics

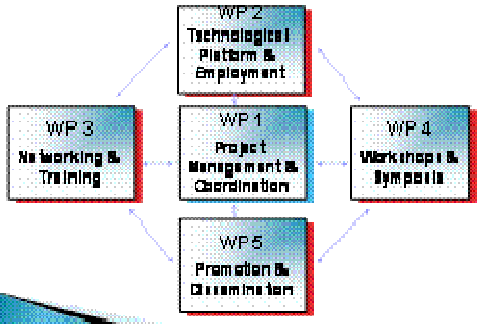
MACPROGEN Project

- National Reference Centre for Genomics and Proteomics
- Project duration 36 months (01.04.2009- 31.03.2012)
- EC contribution: 894.000 Euros
- One beneficiary project – RCGEB, MASA
- EU Partners

EU Partners

- Wilhelm Johannsen Centre for Functional Genome Research, University of Copenhagen, Copenhagen, Denmark
- Hannover Medical School, Gynaecology Research Unit, Hannover, Germany
- University of Barcelona, Faculty of Medicine, Department of Physiological Sciences, Human Genetics Laboratory, Spain
- University of Verona, Department of Mother and Child, Biology and Genetics, Section of Biology and Genetics, Verona, Italy
- University of Copenhagen, Division of Genetics and Bioinformatics, IBHF, Copenhagen, Denmark
- Institute Paoli Calmettes, Molecular Oncology Department, Oncogenomic Group, Marseille, France
- Adaptive Institute of Biomedical Research, Henri Mondor Hospital, Créteil, France
- Institut Cochin, Inserm , University Paris Descartes, Paris, France
- Laboratory for Cytogenetics and Genome Research, Department of Human Genetics, Biomedical Sciences Group, K.U. Leuven Leuven, Belgium

MACPROGEN Workpackages



Main research interest during the MACPROGEN project

- Technology
  - Microarray
  - 2D-DIGE
- Research fields of special interest
  - Cancer genetics (breast cancer)
  - Male infertility
  - Globin genes
  - Deafness
  - Early and non-invasive prenatal diagnosis

Molecular Profiling of Male infertility

- Project (01.01.2010 - 31.12.2012)
- Funded by ICGEB-Trieste
- Research activities:
  - *Screening (AZF rearrang. & chrom. aneupl.)*
  - *Fine mapping of Y deletions*
  - *Role of AZFc partial deletions/duplications*
  - *CFTR and AR genes mutations*
  - *Association studies (polymorph. in differ. genes)*
  - *Spermatozoal RNA profiles*
  - *Proteomic profiling*

Y chromosome structural & functional changes in malignant diseases

- Project (01.10.2010 - 31.09.2012)
- Funded by Ministry of science and education of R. Macedonia
- Research activities:
  - ***Y loss and AZF rearrangements in children at malignancies;***
  - ***Y chromosome haplogroups and haplotypes;***
  - ***TSPY gene copy number in prostate and testicular cancer;***
  - ***Expression studies of several Y chromosome genes;***
  - ***Array CGH and whole genome association studies in males with Y chromosome loss***

Genetics of Deafness

- Project (01.01.2009 - 31.12.2011)
- Funded by MASA
- Research activities:
  - *Mutations in GJB2, GJB6 and GJA1 genes*
  - *Genotype/phenotype correlation in GJB2 patients*
  - *Mitochondrial DNA mutations*
  - *Pendred syndrome (SLC26A4 mutations)*
  - *Array CGH*

Breast cancer research

- Project (01.01.2010 - 31.12.2012)
- Funded by MASA
- Research activities:
  - *BRCA1 and BRCA2 mutations in familial breast cancer*
  - *BRCA1 and BRCA2 polymorphisms*
  - *1100delC mutation in CHEK2 gene*
  - *Association of AR CAG repeats with breast cancer*
  - *CYP2D6 genotyping in patients treated with Tamoxifen*
  - *miRNA profiling of breast cancer tissues*
  - *Identification of cancer specific proteins*

2-D DIGE preliminary experience in RCGEB "Georgi D Efremov"

Katarina Davalieva

2nd Genomics and Proteomics Workshop
June 27-29, 2011

Proteomic analysis of infiltrating ductal carcinoma tissues by coupled 2-D DIGE/MS/analysis

2-D DIGE Katarina Davalieva June, 2011

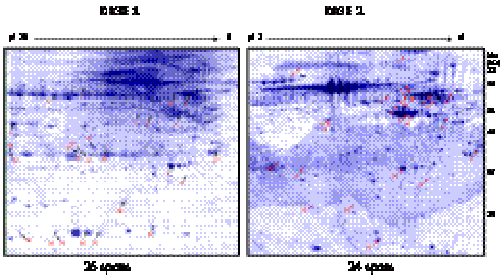
2D-DIGE of breast cancer - experimental design

- Objectives of the study
  To identify specific proteins in infiltrating ductal carcinomas, whose expression is distorted compared to normal breast tissue.
- Samples
  3 tumor and 3 non-tumor tissues from different individuals
- Patients selection
  - Invasive Ductal Carcinoma - Stage IIA
  - ER/PR positive
  - Her 2 positive
- Sample preparation
  - Liquid nitrogen
  - 2D-Clean-Up Kit (GE Healthcare)
- IEF - IPG strips pH 3-10 and pH 4-7, 24 cm
- SDS-PAGE - 12.5% acrylamide gel

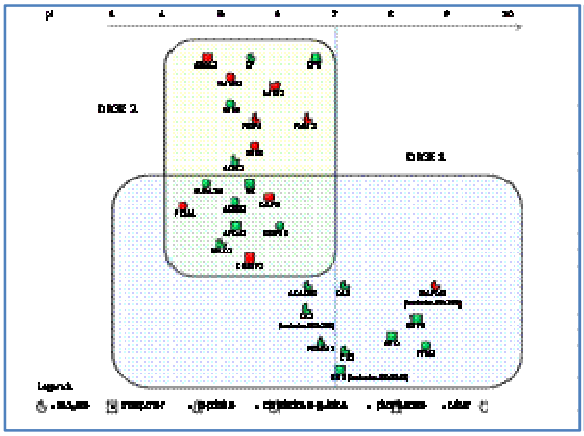
2D-DIGE Katarina Davalieva June, 2011

Proteins of interest

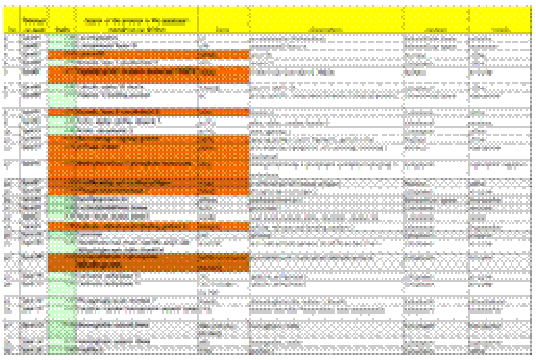
- Anova < 0.05
- Ratio > 2



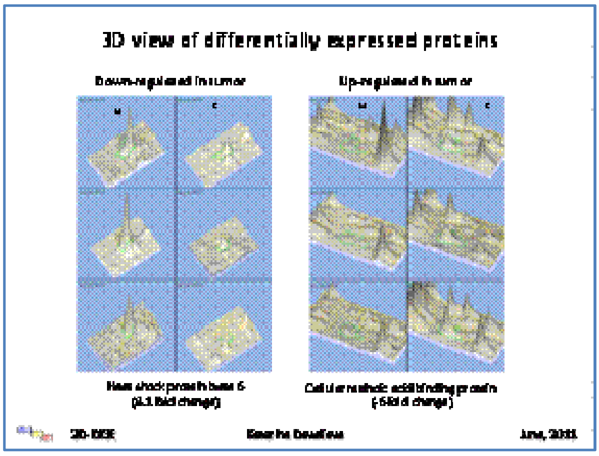
2D-DIGE Katarina Davalieva June, 2011



Final results

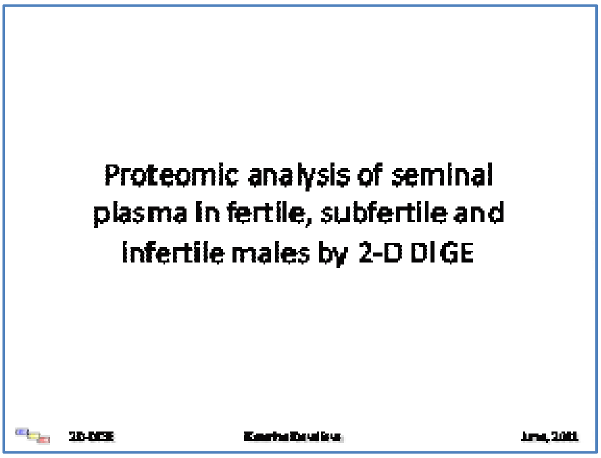


2D-DIGE Katarina Davalieva June, 2011



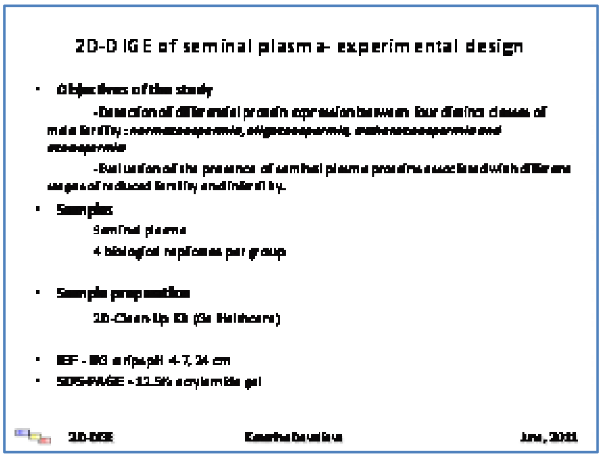
3D view of differentially expressed proteins

2D-DIGE Katarina Davalieva June, 2011



Proteomic analysis of seminal plasma in fertile, subfertile and infertile males by 2-D DIGE

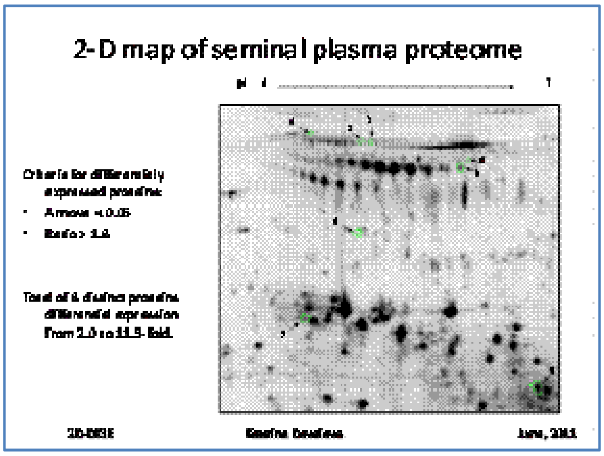
2D-DIGE Katarina Davalieva June, 2011



2D-DIGE of seminal plasma - experimental design

- Objectives of the study
  - -Detection of differential protein expression between four distinct classes of male fertility: normozoospermia, oligozoospermia, asthenozoospermia and azoospermia
  - -Evaluation of the presence of seminal plasma proteins associated with different stages of reduced fertility and infertility.
- Samples
  - Seminal plasma
  - 4 biological replicates per group
- Sample preparation
  - 2D-Clean-Up Kit (GE Healthcare)
- IEF - IPG strips pH 4-7, 24 cm
- SDS-PAGE - 12.5% acrylamide gel

2D-DIGE Katarina Davalieva June, 2011

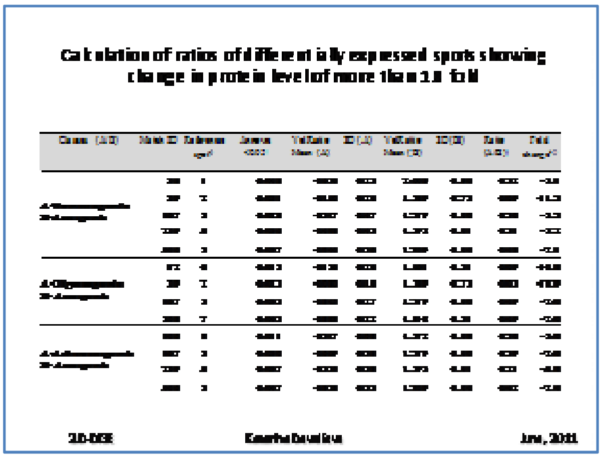


2-D map of seminal plasma proteome

Criteria for differentially expressed proteins:
- Anova < 0.05
- Ratio > 1.8

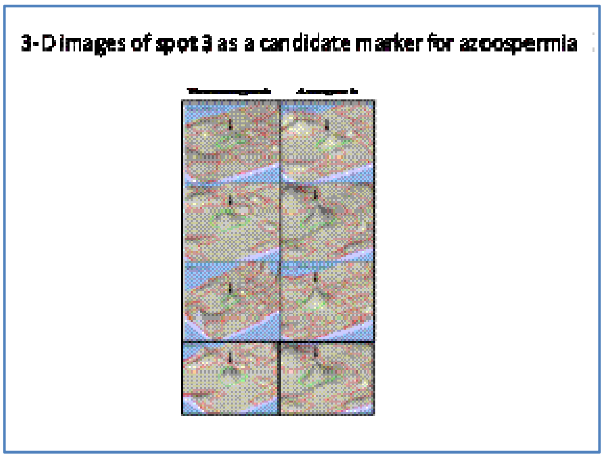
Total of 6 distinct proteins differential expression From 2.0 to 11.5-fold.

2D-DIGE Katarina Davalieva June, 2011



Calculation of ratios of differentially expressed spots showing change in protein level of more than 1.8 fold

2D-DIGE Katarina Davalieva June, 2011



3-D images of spot 3 as a candidate marker for azoospermia

Non-coding RNA profiling by RNA-Seq

Lyubomira Chakalova

ncRNAs fulfill a variety of functions

Gene silencing and activation in *cis* and *trans*

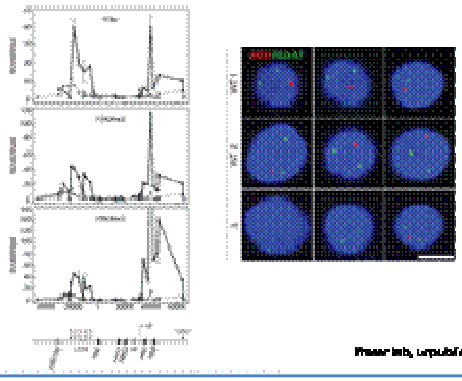
Genomic reprogramming

Enhancer activity

Nucleators of nuclear structures such as paraspecles

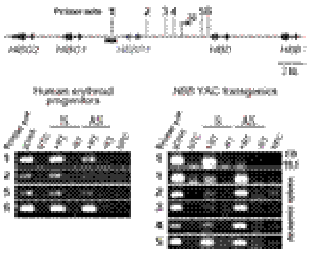
Many genomic loci are riddled with non-coding RNAs of unknown functions

One example: the *HBB* locus



Fraser lab, unpublished

Tip of the iceberg...



Fraser lab, unpublished

ncRNA discovery and characterisation

We need ways to determine:

- sizes
- transcription start sites
- direction of transcription
- half-life
- polyadenylation status

RNA-Seq?

Library preparation

Adapting the Cloonan method for Illumina

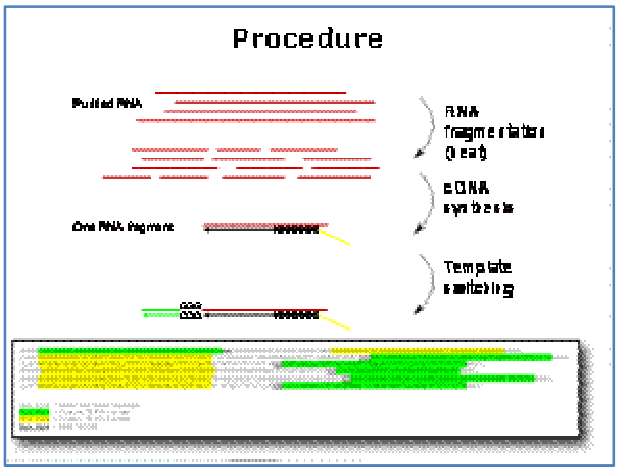
Published procedure by Cloonan *et al*, *Nat Methods*, 2008: compatible with SOLiD sequencing

Illumina oligonucleotides (Cloonan, *et al*., *Nature*, 2008):

Random primer with Illumina adapter

Template switching primer with Illumina adapter

Illumina PE PCR primers


Procedure

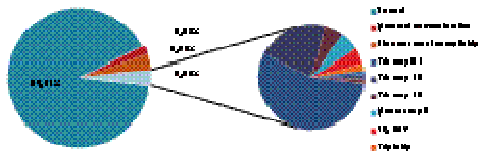
Early and Non-Invasive Molecular Prenatal Diagnostics

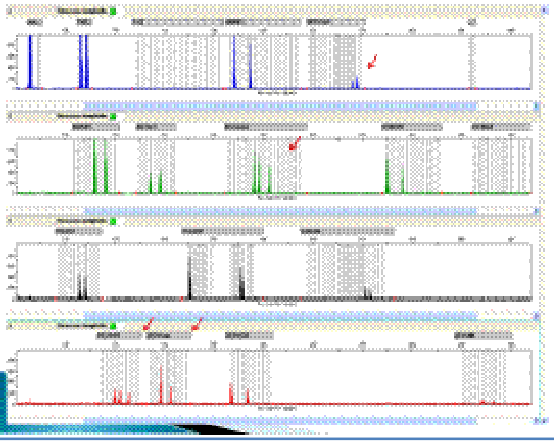
Svetlana Madjunkova
RCGEB, Skopje
28 June 2011

Background: chromosomal aneuploidies

Rapid prenatal diagnosis for chomosomal aneuploidies:

- Multiplex QF-PCR testing highly polymorphic STRs on chr 13, 18, 21, X and Y





Background: monogenic diseases

- β-Thalassaemia
- Cystic fibrosis
- Spinal muscular atrophy
- Duchenne/Becker muscular dystrophy
- Hemophilia A
- Lech Nichan syndrome
- Rett syndrome
- Phenylketonuria
- Galactosemia
- Pseudohipoaldosteronizam

Future project

- Early prenatal diagnosis
- Non-invasive prenatal diagnosis using cell-free fetal DNA from maternal plasma
- Aneuploidy related biomarker discovery by 2D-DIGE

Early prenatal molecular diagnosis

- Screening subtelomeric regions by MLPA
- Screening for common microdeletion/duplication syndromes by MLPA
- Implementing aCGH

Non-invasive prenatal diagnosis using cell-free fetal DNA from maternal plasma

- Non invasive determination of paternally inherited traits
  - fetal sex (Y specific loci) ([illegible])
    - X-linked disorders (Duchenne/Becker muscular dystrophy, Hemophilia A)
  - fetal RhD status in RhD (-) mothers
    - RHD fetal-maternal incompatibility
  - fetal specific mutant alleles
    - monogenic diseases (β-thalassemia, cystic fibrosis)
- Non invasive diagnosis of chromosomal aneuploidies
  - differentialy methylated fetal DNA markers

Aneuploidy related biomarker discovery by 2D-DIGE

- Identification of aneuploidy specific protein biomarkers
  - In maternal plasma
  - In placenta

Collection of samples

- Creating comprehensive patient database
- Types of samples
  - Chorionic villi
  - Amniocytes
  - Whole blood from both parents
  - Maternal plasma
  - Placentas

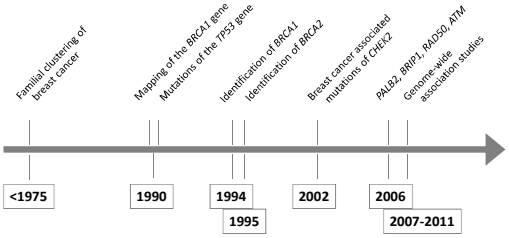
We are open for collaboration

Molecular genetics of breast cancer

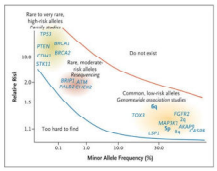
Thilo Dörk-Bousset

Hannover Medical School, Germany

From rare germ-line mutations with high penetrance to common gene variants with low penetrance



Methods to identify breast cancer susceptibility alleles in relation to their frequency and penetrance



taken from: Foulkes W., *N England J Med* 359(20): 2143-2153

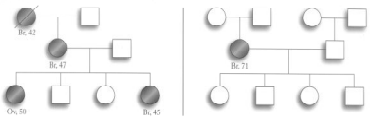
- high penetrance, rare: Segregation analysis in „multiple-case" families
- intermediate penetrance, rare: Sequencing of candidate genes
- low penetrance, common: Case-control association studies

Breast cancer susceptibility genes (I): Moderate to high penetrance (relative risks 2- 20)

Gene	Locus	Associated disease	Cases
BRCA1	17q21	Fam. Breast- and Ovarian Cancer	2-3 %
BRCA2	13q12	Fam. Breast- and Ovarian Cancer	1-2 %
PALB2	16p12	Fam. Breast-, Ovarian and Gastric Cancer	0.5- 1 %
TP53	17p13	Li-Fraumeni Syndrome	very rare
MSH2	2p22	Muir-Torre-Syndrome	very rare
STK11	19p13	Peutz-Jeghers-Syndrome	very rare
PTEN	10q23	Cowden´s Disease	very rare
CDH1	16q22	Fam. lobular Breast Cancer	very rare
CHEK2	22q13	Hereditary Breast Cancer	~ 3 %
ATM	11q23	Hereditary Breast Cancer	~ 2 %
NBN	8q21	Hereditary Breast Cancer	< 1 %
MRE11A	11q21	Hereditary Breast Cancer	< 1 %
RAD50	5q31	Hereditary Breast Cancer	< 1 %
RAD51C	17q25	Hereditary Breast Cancer	< 1 %
BRIP1	17q23	Hereditary Breast Cancer	< 1 %

Familial breast cancer has a strong hereditary component

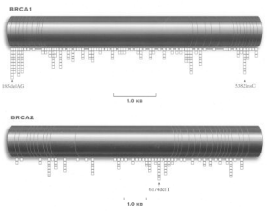
„familial" (*germline mutation ?*) | „sporadic" (*somatic mutation ?*)



Some criteria for the analysis of the *BRCA1/ BRCA2* genes in breast cancer cases (recommendations for German patients):

* three or more females affected by breast cancer
* two or more females with breast and ovarian cancer
* two females with breast cancer, one of them by age < 50 ys
* one female with bilateral breast cancer and her primary < 40 ys
* one female with breast cancer < 30 ys
* one female with ovarian cancer < 40 ys

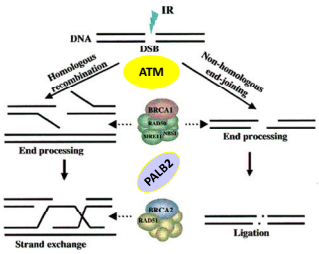
Allelic heterogeneity of *BRCA1* and *BRCA2* mutations



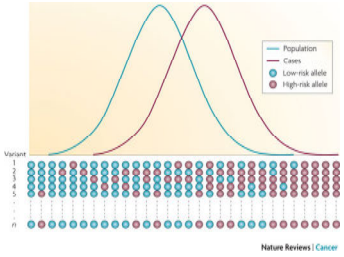
Common mutations:	BRCA1	BRCA2
Caucasians	5382insC Cys61Gly 4184del4	3034del4 9326insA 5950delCT
African-Americans	943ins10 1832del5 5296del4	2816insA
Asians	589delCT Leu63Ter	4265delCT
Ashkenazim	185delAG	6174delT
Philippines		4859delA
Netherlands	2804delAA del(e13) del(e22)	
Iceland		999del5

BRCA1 and BRCA2 mutations are distributed throughout the whole gene, mainly of the truncating type, and they are found in all populations tested.

BRCA1 and BRCA2 normally function in concert with other gene products in the repair of DNA double-strand breaks



Hypothesis: Combinations of multiple common risk alleles make a significant contribution to the overall breast cancer risk



First genome-wide association study for breast cancer

Stage 1: 205.586 SNPs, 390 patients/ 364 controls
Stage 2: 9.250 SNPs, 3990 patients/ 3916 controls
Stage 3: 30 SNPs, 22656 patients/ 20973 controls

Locus 1: *FGFR2*	10q	per allele OR 1,26	p= $2x10^{-76}$
Locus 2: *MAP3K1*	5q	per allele OR 1,13	p= $7x10^{-20}$
Locus 3: *TOX3*	16q	per allele OR 1,11	p= $2x10^{-19}$
Locus 4: *MYC*	8q	per allele OR 1,08	p= $5x10^{-12}$
Locus 5: *LSP1*	11p	per allele OR 1,07	p= $3x10^{-9}$

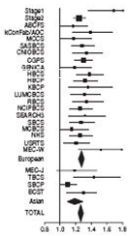
(Easton et al.; *Nature* 2007)

Breast cancer susceptibility genes (II): Low penetrance (relative risks 1.1- 1.5)

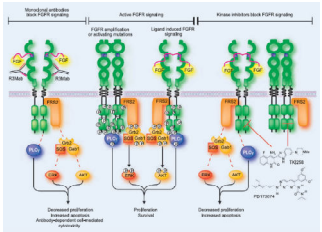
Chromosome	SNP	Candidate gene	Frequency
10q	rs2981582	*FGFR2*	0,38
16q	rs12443621	*TOX3*	0,46
5q	rs889312	*MAP3K1*	0,28
11p	rs3817198	*LSP1*	0,30
8q	rs13281615	-	0,40
2p	rs1045485	*CASP8*	0,87
2q	rs13387042	-	0,54
5p	rs4415084	*MRP530*	0,40
6q	rs2180341	*RNF146*	0,21
3p	rs4973768	*SLC4A7/ NEK10*	0,46
17q	rs6504950	*COX11*	0,73
1p	rs11249433	-	0,39
14q	rs999737	*RAD51L1*	0,76
6q	rs2046210	*ESR1*	0.34

...**many** more coming soon...

New possibilities towards a „targeted therapy“ for breast cancer

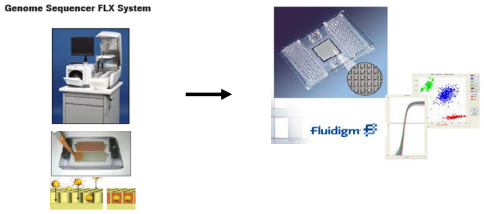


Easton DF, et al. 2007 Nature 447(7148): 1087-1093.



Hynes NE, Dey JH. 2010. Potential for Targeting the Fibroblast Growth Factor Receptors in Breast Cancer. Cancer Res 2010;70 5199-5202

Probable next steps:

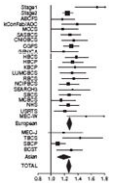


Deep genomic sequencing to identify the (rarer ?) causal variants

High-throughput genotyping to screen large case-control studies for the causal variants

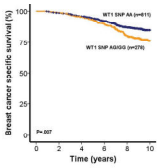
In search of genetic biomarkers for breast cancer prognosis

Risk factor: ***FGFR2***
relative frequency in patients *versus* controls (Odds ratio)



Easton DF, et al. 2007
Nature 447(7148): 1087-1093.

Prognostic factor: ***WT1***
breast cancer specific survival in patients with lobular breast cancer (n= 1089)



Heuser et al. 2011, submitted.

„5-10% of all breast cancers are genetically determined"

o.k, I mean: „5-10% of all breast cancers are genetically determined by inherited factors"

Fact: Some 5-10% of breast cancer patients have a strong *family history* of the disease.

But what is the true fraction of breast cancers that are – at least in part – genetically influenced ?

Yesterday: ~ 5-10% of patients

Today: probably all

Summary

- Sequencing efforts and large-scale case-control studies have identified several new gene mutations and genetic variants that contribute to breast cancer.
- Moderate- to high-penetrance mutations are rare and are frequently restricted to founder populations. Low-penetrance variants can be common and shared between many different populations.
- Identification of breast cancer susceptibility alleles can pave the way for:
  - Elucidating causes and biological mechanisms of breast cancer
  - Stratification of healthy individuals by genetic risk
  - Stratification of breast cancer patients for a better therapy

The Gynaecology Research Unit at Hannover Medical School
(Head of the Clinics: Professor Peter Hillemanns)

Present Team: Natalia Bogdanova, Natalia Dubrowinskaja, Rosa Landwehr, Britta Wieland, Peter Schürmann, Bianca Heurich, Thilo Dörk-Bousset

(www.mh-hannover.de/molgynres.html)

Research Partners on Breast Cancer:

* Natalia Antonenkova (Minsk, Belarus)
* Elza Khusnutdinova (Ufa, Russia)
* Evgeny Imyanitov (St. Petersburg, Russia)
* Jan Lubinski (Szczecin, Poland)
* Georgia Chenevix-Trench (Brisbane, Australia)
* Paul Pharoah, Douglas Easton (Cambridge, UK)

The Breast Cancer Association Consortium is supported by the European Union through the COST Action BM0606 and the FP7 project „Collaborative Oncological Genetic Studies"

University of Plymouth

safenetwork

Using DiGE to define markers for prenatal diagnosis (and screening)

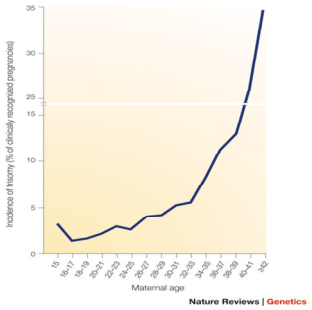
2nd Genomics and Proteomics workshop, Skopje, Macedonia

June 28th 2011

Neil Avent

safenetwork

Increase of fetal aneuploidies

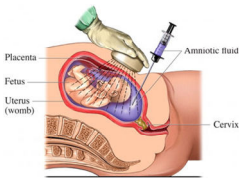


Risk = 1/250 aneuploidy

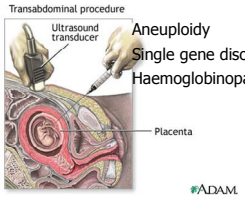
Risk = 1/100 fetal loss by AC / CVS

Men – loss in fertility

Current prenatal diagnosis requires invasive procedures



Amniocentesis



Chorionic villus sampling

Aneuploidy
Single gene disorders
Haemoglobinopathies

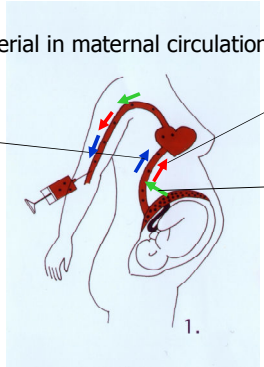
1% risk of miscarriage

Not possible before 11 weeks'

University of Plymouth

safenetwork

Fetal Material in maternal circulation



Fetal cells in the Maternal circulation

Free Fetal DNA in Maternal plasma

Free Fetal RNA (From Placenta)

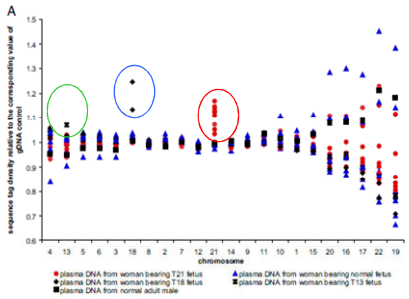
Fetal Proteins in Maternal plasma (Placenta)

www.safenoe.org

University of Plymouth

Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L. & Quake, S.R. (2008) Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci U S A*, ***105, 16266-16271***

NGS- It works for NIPD but with huge cost implications and time



safenetwork

Better Down plasma biomarkers ?

- New generation screening tests
- Replace those currently used? AFP, β-HCG, PAPP-A, Inhibin-A
- 99%+ accurate?
- Would genetic diagnosis then be necessary?
- Almost certainly for confirmation

Non-Invasive Analysis of Serum markers in Mothers Blood

Triple Serum Protein Test

SCREENING:	AFP:	Estriol:	hCG:
Down Syndrome	Low	Low	High
Trisomy 13	Norm	No Data	Low
Trisomy 18	Low	Low	Low
Open Neural Tube Defects	High	Normal	Normal
IUGR, preterm birth, stillbirth	High	No Data	No Data
Multiple Gestation	High	Normal	High

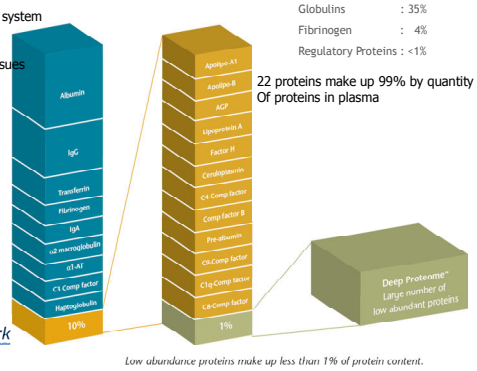
Indicative but not diagnostic…

http://library.med.utah.edu/WebPath/TUTORIAL/PRENATAL/PRENATAL.html

Plasma Proteomics: High complexity

Plasma most complex proteome-
Coagulation cascade
Complement
Humeral immune system
Cytokines

Plus bathes all tissues



Albumins : 60%
Globulins : 35%
Fibrinogen : 4%
Regulatory Proteins : <1%

22 proteins make up 99% by quantity
Of proteins in plasma

Low abundance proteins make up less than 1% of protein content.

Biomarker discovery

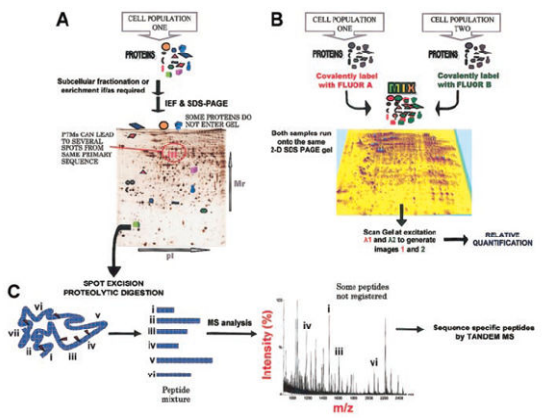
- Major focus of Post genome sequence activity
- Protein markers of disease for diagnostics
- Plasma proteins major consideration

WP3B SAFE Network of excellence activity – Biomarker discovery for :

- Aneuploidy
- Pre-term labour
- Preeclampsia

Proteomics work flow

- Protein isolation
- Protein separation
- Protein Identification – staining
- Protein Identification – Excision and trypsin digestion
- Protein Identification – Mass spectrometry



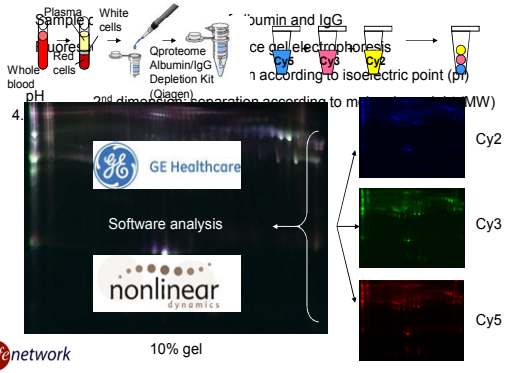
SAFE Plasma Proteomics
(and transcriptomics)

Dr Lyn Chitty
Dr Wendy Heywood
Dr Kevin Mills

Prof Neil Avent
Dr Tracey Madgett
Dr Debbie Maddocks

Prof Peter Soothill
Kin Choi

Sample Depletion and DIGE protocol

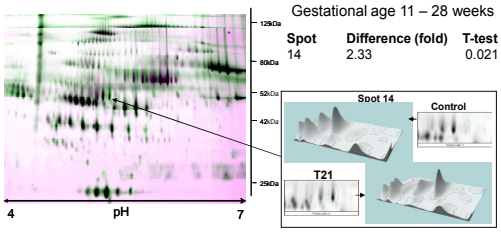


10% gel

safenetwork

T21 samples:

- 6 pairs of age- and sex- matched control and T21 samples
- Study performed at UCL and UWE



safenetwork

T21 samples: pH ranges

- UCL – 1st trimester (<14 weeks)
- UWE – 2nd trimester (14-32 weeks)

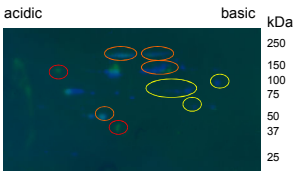
- pH 4.5 – 5.5
- pH 5.3 – 6.5
- pH 6.0 – 9.0

- In total, 11 to 14 pairs of control/T21 maternal plasma samples per pH range and trimester
- Lots of candidate biomarkers identified!

safenetwork

Results: 2D Western Blotting

Polyclonal antibody to protein of interest – possibly protein of interest is bound to other proteins



Key:

Red channel = non-pregnant

Blue channel = control

Green channel = T21

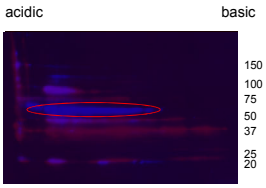
present in [illegible] and T21

Western blot films have been scanned and then converted to single channel images (RGB) using Adobe Photoshop

safenetwork Apologies – Identity of protein species not disclosed - IP

Results: 2D Western Blotting

Polyclonal antibody to protein of interest – possibly protein of interest is bound to other proteins



Key:

Red channel = non-pregnant

Blue channel = control

Green channel = T21

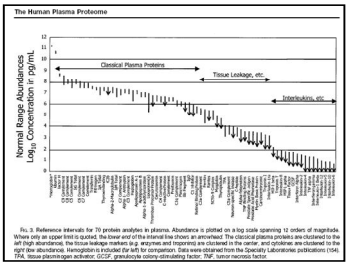
present in [illegible] in non-pregnant

Western blot films have been scanned and then converted to single channel images (RGB) using Adobe Photoshop

safenetwork

Both Upregulated and Downregulated protein IDs made :

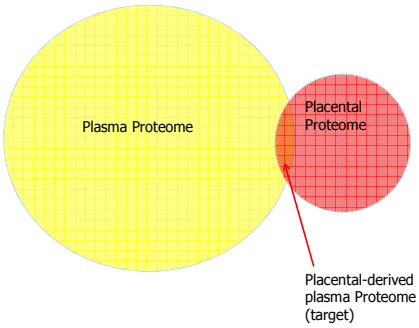
- Complement proteins
- Annexins
- Many proteins already described in the literature
- Mostly high abundance proteins
- Rarely >1.5 fold up- or down-regulation
- Significantly – almost all differences are down to differences in post-translational modifications

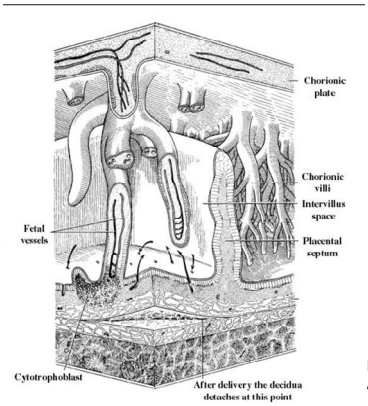


The Plasma Proteome :

- Most complex
- Potential for the complete human proteome due to tissue leakage
- Tissue leakage – markers of disease (cancer & myocardial infarct)
- New proteomics technologies directed to low-abundance species

Biomarker discovery – A question of scale

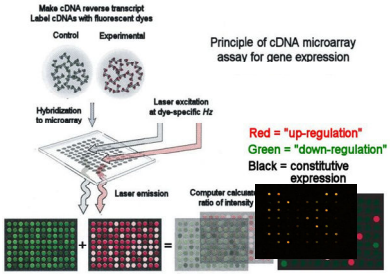




Placental Anatomy

safenetwork
Special Advances in Fetal Evaluation

Normal Trisomy 21



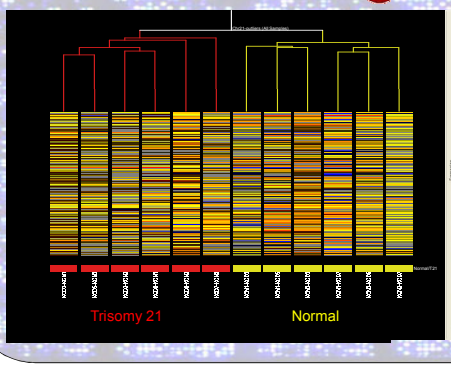
Principle of cDNA microarray assay for gene expression

Red = "up-regulation"
Green = "down-regulation"
Black = constitutive expression

Hierarchical clustering – Gene condition tree

Principle Component Analysis

safenetwork

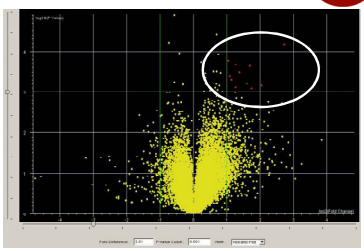


safenetwork
Special Advances in Fetal Evaluation

University of BRISTOL

Kin Choi
Prof Peter Soothill

UWE- Dr Debbie Maddocks



Placental Transcriptomics – Affymetrix U0133 arrays
11 genes with p>0.0001
3 1st trimester placentas Normal; 3 1st trimester placentas, Downs

Larger Microarray study

In collaboration with Prof Kypros Nicolaides and Dr Ranjit Akolekar, King's college London

20 Trisomy 21 CVS and 20 age-matched controls with plasma samples

Agilent whole genome arrays

Large number of new biomarkers under validation (lack of antibodies to 90%)

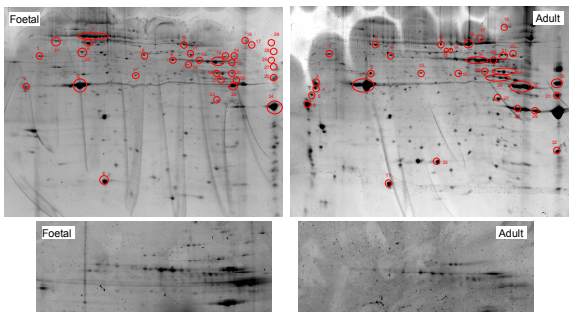
safenetwork
Special Advances in Fetal Evaluation

Proteomics of fetal erythroblasts

- Funded by National Blood service and SAFER Strep FP6 project (SAFE spin-out project)
- Identification of fetal-erythroblast specific markers
- Comparison of cordocentesis and adult erythroid cells
- Provide better targets for fetal cell isolation

Protein Mapping & Marker Discovery

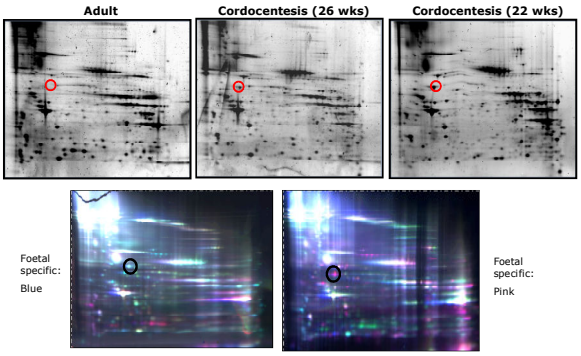
2D Electrophoresis (2DE): Erythroid & Erythroblast membranes



Analysed by PDQuest computer software – Many up and down regulated foetal proteins – 21 patented (Submitted, August 2007)

Marker Specificity – Hsp60

2D Electrophoresis (2DE) and Fluorescent 2DE of erythroid membranes

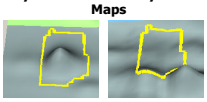


Foetal Vs Adult

Marker specificity – Hsp60

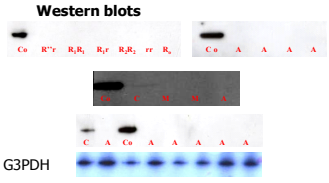
2DE software analysis, western blotting and immunofluorescence

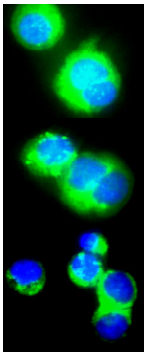
DeCyder Software analysis of 2DE Maps



Foetal **Adult**

Western blots





Analysis of Hsp60 cellular localisation in foetal cultured erythroblasts

Hsp60: Green (FITC)

Nucleus: Blue

(Alexa 488)

safenetwork
Special Advances in Fetal Evaluation

Fetal erythroblast markers

- New generation fetal erythroid markers determined, and patent submitted
- Some expressed on other peripheral blood mononuclear cells
- Requires parallel development of microfluidics (SAFER project objective)
- Requires further funding (thanks Wellcome trust!)

safenetwork
Special Advances in Fetal Evaluation

Conclusions

Plasma proteomics- huge logistic exercise to identify candidate biomarkers (Placental origins)

Transcriptomics of target tissue (Placenta) more productive approach?

New candidate biomarkers being validated

New markers of fetal erythroblasts may improve fetal cell isolation techniques?

Acknowledgements

University of Bristol – Prof Peter Soothill, Kin Choi
UCL- Prof Lyn Chitty, Dr Wendy Heywood
University of Basel – Prof Sinuhe Hahn
URV Tarragona – Prof Ciara O'Sullivan

safenetwork
Special Advances in Fetal Evaluation

KATHOLIEKE UNIVERSITEIT LEUVEN

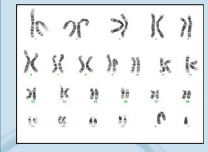
GENOMIC ARRAYS IN DIAGNOSTICS

Paul Brady

MACPROGEN
Macedonia
28 June 2011

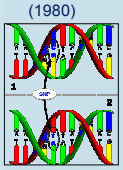
Historical Perspective : The Genetic Causes of Our Differences

Chromosomes (1960)



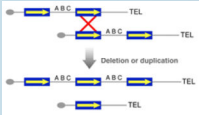
Variants are rare

Single nucleotide polymorphisms (SNPs) (1980)



Frequent:
- 1 SNP every 1000 bp
- 0.1% difference between 2 individuals
- 3 Mb difference

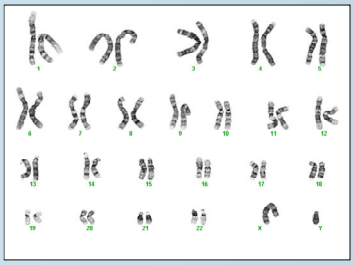
Copy Number Variations (CNVs) (2004)



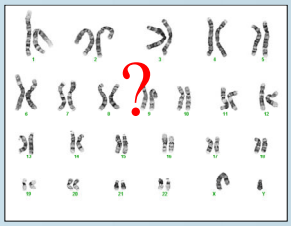
Very frequent:
- 1000 CNVs/2 individuals
- 0.7% of the genome is copy variable between 2 individuals
- 21 Mb difference!

Karyotyping – Genome-Wide Screening for Chromosome Anomalies

Main Limitations;
- SIZE
- Time-Consuming
- Requires Expertise
- Subjective
- Not Molecular



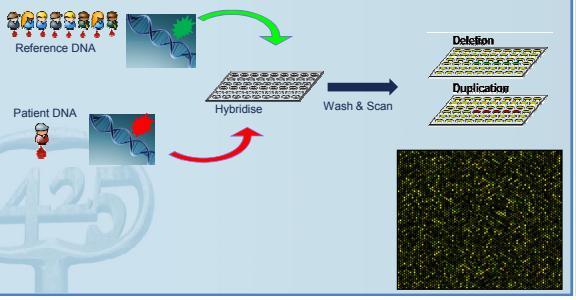
For The Majority of MCA / ID Patients There Is No Diagnosis



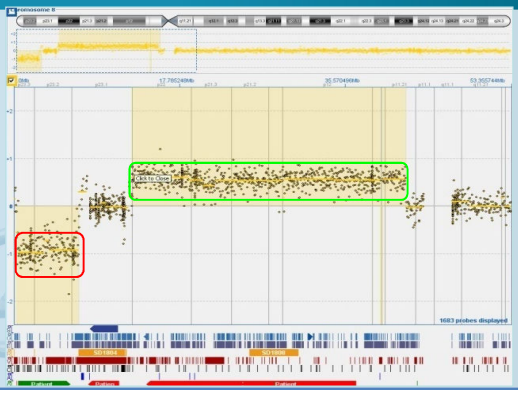
Question : Can Submicroscopic Imbalances Explain The Cause Of The MCA / ID ?

The Array Revolution

Molecular Karyotyping Using DNA Microarrays

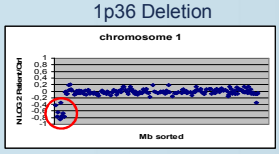


Detection Of Genomic Gains & Losses



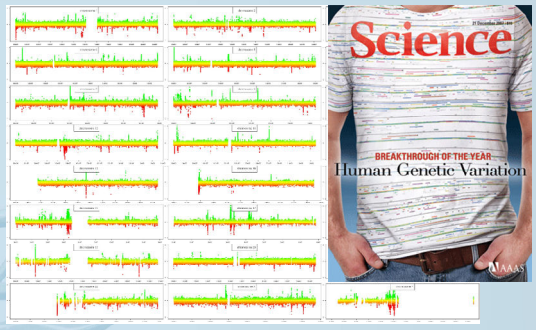
Arrays Reveal High Incidence Of Submicroscopic Imbalances

1p36 Deletion



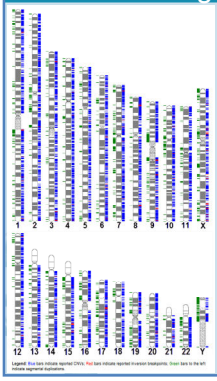
Improvement Of Detection Rates
- Karyotyping <3%
- Sub-telomere FISH/MLPA ~2.5%
- 1Mb BAC array ~10%
- 180K oligonucleotide array ~17%

The Bad News : We Are All Copy Variable



~35% of the Genome is Copy Variable in Normal Individuals

Databases of Genomic Variants : Catalogue of 'Benign' CNVs



- Databases Of 'Benign' CNVs Have Limited Value For Clinical Assessment
- Beware of 'HapMap bias'

Database of Genomic Variants
Genome-wide view of CNVs

Summary Statistics

Total entries: 101923 (hg18)
CNVs: 66741
Inversions: 953
InDels (100bp-1Kb): 34229
Total CNV loci: 15963
Articles cited: 42

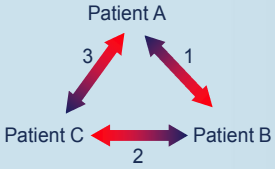
Last updated: Nov 02, 2010
Join our mailing list

Toronto Database of Genomic Variants

Reference DNA

- DNA From Normal Individual
  - Who is Normal?
- DNA From A Mixture Of Individuals
  - How many?
  - Which?
- DNA From Other Patients
  - When?
  - Three Way Hybridisations
- DNA From The Same Individual
  - For acquired disorders only

Loop Design



Cy5 vs Cy3
1. Patient A vs Patient B
2. Patient B vs Patient C
3. Patient C vs Patient A

Patient A = Array 1 (confirmed in 3)
Patient B = Array 2 (confirmed in 1)
Patient C = Array 3 (confirmed in 2)

ONLY PATIENTS WITH DIFFERENT PHENOTYPES !!

BMC Bioinformatics

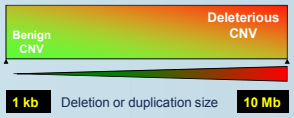
Methodology article — Open Access

An experimental loop design for the detection of constitutional chromosomal aberrations by array CGH

Joke Allemeersch[1], Steven Van Vooren[2], Femke Hannes[3], Bart De Moor[2], Joris Robert Vermeesch[3] and Yves Moreau[*1,2]

Address: [1]Microarray Facility, VIB, Leuven, Belgium, [2]ESAT-SISTA, KU Leuven, Leuven, Belgium and [3]Center for Human Genetics, University Hospital Gasthuisberg, Leuven, Belgium

Clinical Validity : What Is The Clinical Significance Of Imbalance?



Cytogenet Genome Res 115:225–230 (2006)
DOI: 10.1159/000095916
Cytogenetic and Genome Research

Molecular karyotyping of patients with MCA/MR: the blurred boundary between normal and pathogenic variation

J.J.L. de Ravel, L. Balikova, B. Thienpont, F. Hannes, N. Maas, J.-P. Fryns, K. Devriendt, J.R. Vermeesch

Centre for Human Genetics, UZ Gasthuisberg, Leuven (Belgium)

With Ever Increasing Resolution, The Boundary Between Benign & Pathogenic CNVs Becomes Blurred!

Choice Of Platform

- CNV array / SNP array or both?
- Depends on Purpose
- Genome-Wide Coverage With Higher Density Coverage of Clinically Significant Targets
- Recommended resolution for constitutional diagnosis is at least 400-500 kb
- Allows for detection of all known recurrent microdeletion / microduplication syndromes

Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies

The Challenge : Which Imbalances Are Causal For The Phenotype?

Conventional Wisdom:

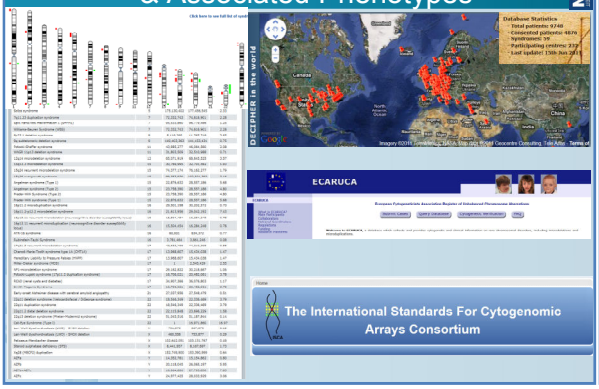
Recurrent imbalances with same phenotype are causal

The larger the size, the more likely causal

Inherited imbalances are benign whilst *de novo* imbalances are causal

Population embedded CNVs are benign

Identification of Recurrent Imbalances & Associated Phenotypes



The Challenge : Which Imbalances Are Causal For The Phenotype?

Conventional Wisdom:

Recurrent imbalances with same phenotype are causal

The larger the size, the more likely causal

Inherited imbalances are benign whilst *de novo* imbalances are causal

Population embedded CNVs are benign

Rare CNVs Megabases in Size Are Observed in Normal Individuals

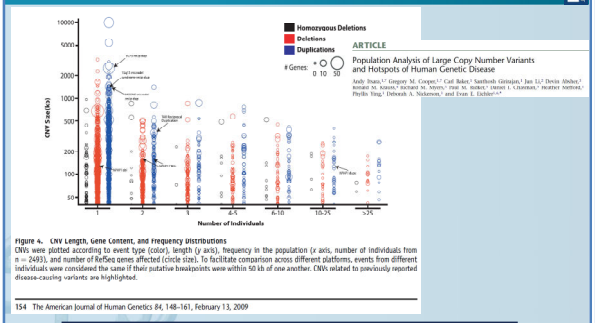


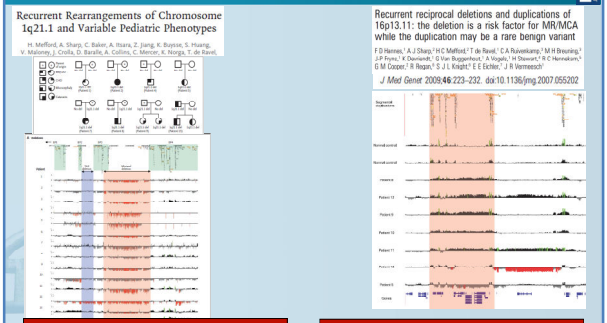
Figure 4. CNV Length, Gene Content, and Frequency Distributions

CNVs were plotted according to event type (color), length (y axis), frequency in the population (x axis, number of individuals from n = 2693), and number of RefSeq genes affected (circle size). To facilitate comparison across different platforms, events from different individuals were considered the same if their putative breakpoints were within 50 kb of one another. CNVs related to previously reported disease-causing variants are highlighted.

154 The American Journal of Human Genetics 84, 148–161, February 13, 2009

Size Alone Is Not A Good Determinant
Nor Occurrence In Apparently Normal Individuals

CNVs As Risk Factors For ID/MCA

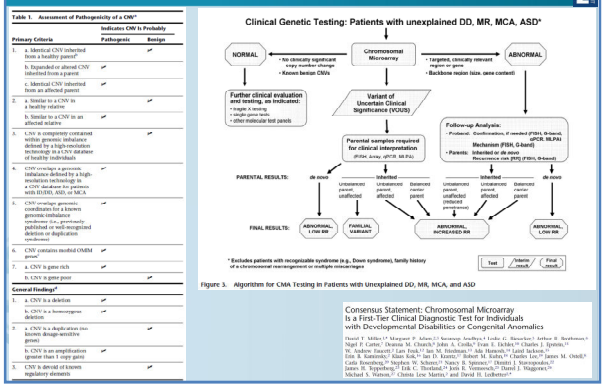


Deletion	Duplication
25/5218 patients	9/5218 patients
0/4737 controls	1/4737 controls
P = 1.1×10^{-7}	P = 0.02

Deletion	Duplication
5/1026 patients	5/1026 patients
0/2014 controls	5/1682 controls
P =0.0048	No Difference

Inherited Does Not Necessarily Mean Benign

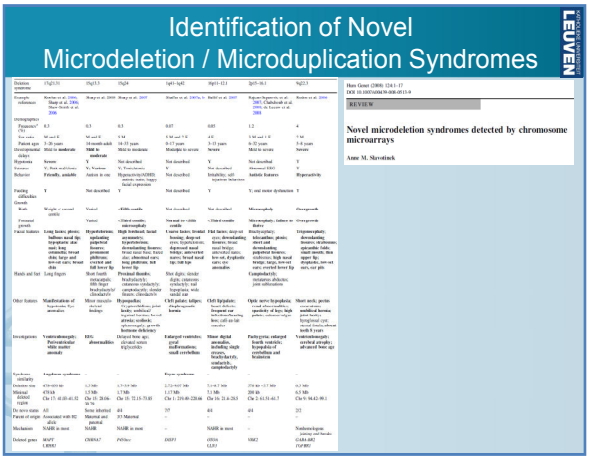
Criteria For Determining Pathogenicity



Databases

The Vast Amount Of Information Requires Bioinformatics Tools To Aid Clinical Interpretation & For Storage Of Data

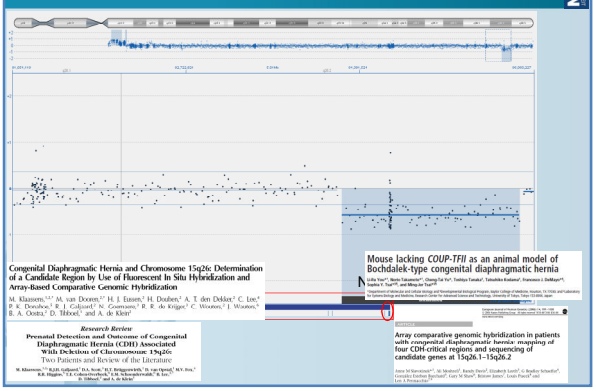
Identification of Novel Microdeletion / Microduplication Syndromes



Refinement Of Causal Regions

- Collection of patients with similar phenotypes and overlapping genomic imbalances
- MDR (Minimal Deleted Region)
- SRO (Smallest Region of Overlap)
- Genotype – Phenotype Correlations are aided by Overlapping Non-Recurrent Imbalances
- Identification of Causal Genes

Refinement of Minimal Deleted Region on 15q26

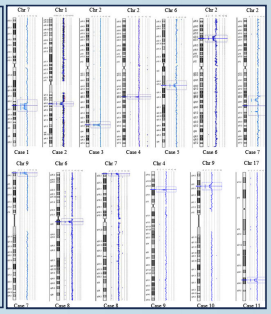


Detection of Mosaics :
Tetrasomy 5p (20%) / Trisomy 21 (40%)

Apparently Balanced Translocations : The Majority Are Unbalanced !

59 cases investigated;

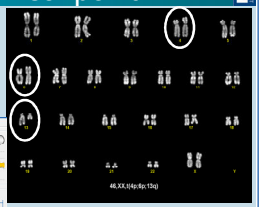
- 41 apparently balanced translocations:
  - 27 patients
  - 40% (11/27) unbalanced
    - 22% (6/27) with deletions at the translocation breakpoints
    - 18% (5/27) with complex rearrangements
  - 14 fetuses: all normal
- 18 complex rearrangements:
  - 13 patients
  - 3 fetuses
  - 2 females with repeated abortions
  - 16/18 (89%) unbalanced

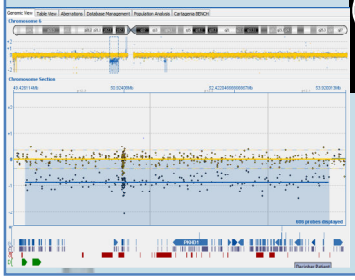


De Gregori, *et al.*, *J. Med. Gen.* (2007)

Submicroscopic Deletion at Rearrangement Breakpoint

- Complex Familial Rearrangement
- 46,XX,t(4p;6p;13q)





4Mb deletion at 6p12 breakpoint

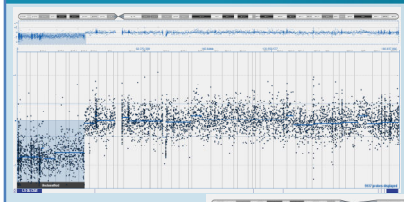
When To Karyotype or FISH?

Array Profiles With Characteristic Signatures Of Unbalanced Rearrangements

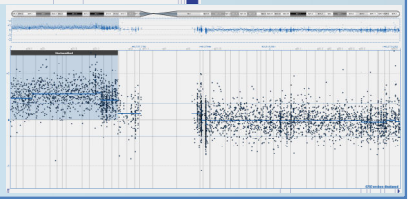
- Unbalanced Translocations; Terminal Del & Terminal Dup
- Inv Dup Del; Terminal Deletion & Interstitial Dup of p/q arm
- Ring Chromosomes; Terminal Del of p & q arm

- Advise parental karyotype to detect balanced rearrangements
- FISH provides positional information for duplications – insertional translocations
- False Positives are a problem inherent to arrays with poor quality

Unbalanced Translocations

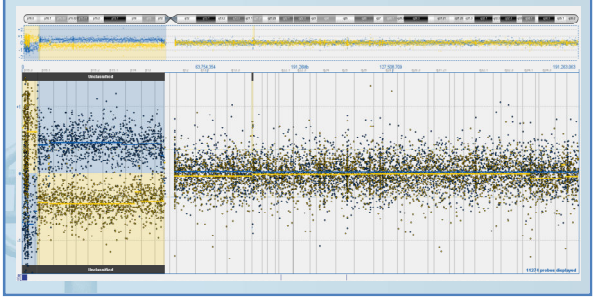


- t(5p;9p)



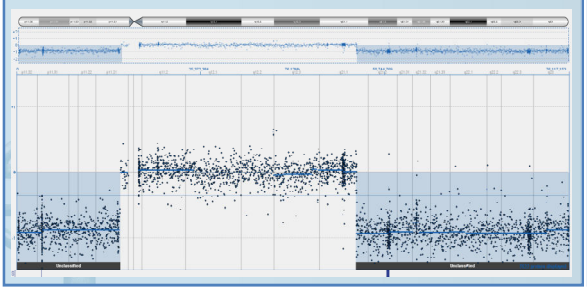
Inverted Duplication Deletion

- 4p inv dup del



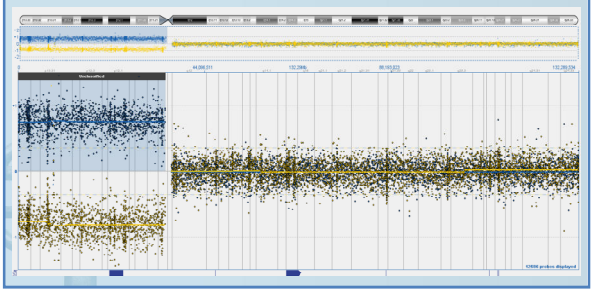
Ring Chromosome

- Ring chromosome 18



Isochromosomes

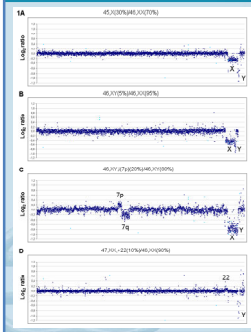
- i(12p) Pallister Killian Syndrome

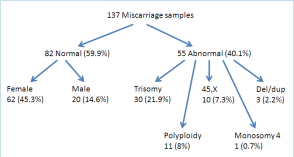


Limitations Of Array CGH

- Truly Balanced Rearrangements **Will Not** Be Detected
- Triploidy – possible to detect 69,XXY with 47,XXY ref
- Our Knowledge of Very Small CNVs is Limited
- Particularly in the Prenatal Setting
- Continuous – Knowledge is Improving
- Static – Reduced Penetrance / Variable Expression
- Imbalances detected by Array CGH represent the additive, not the allele-specific results
- Most Disease is caused by mutations and only a minority of disorders are caused by CNVs

Miscarriages: Array CGH As First Tier Test





Trisomies	
chromosome 2	1
chromosome 3	2
chromosome 7	1
chromosome 8	2
chromosome 9	1
chromosome 13	2
chromosome 15	6
chromosome 16	4
chromosome 18	2
chromosome 19	1
chromosome 20	1
chromosome 21	6
chromosome 22	3

Robberechts et al., Gen. Med.. 2009

Mosaicism can be detected as low as 5% : Sensitivity ≥ Karyotyping

Towards Prenatal Diagnosis

TERRA INCOGNITA

How to deal with
- Variable Expression / Reduced penetrance?
- Unclassified Variants?
- Late Onset Disorders?
- Clinically Relevant Finding Unrelated To Phenotype;
  - In Foetus?
  - In Parents?

What is 'Normal' ?

Who Decides ?

Prenatal Array CGH In Routine Diagnostics

Initial Strategy in Leuven (approved by ethical committee)

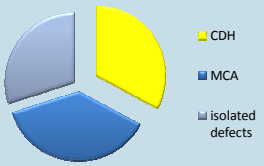
- Only foetuses with abnormal ultrasound findings
- Interpretation by both a cytogeneticist & clinical geneticist
- Report only relevant findings (proven association with MCA)
- No connection between the original data and patients
- Include parents in the analysis

Prenatal Array CGH for US anomalies

Criteria for inclusion;

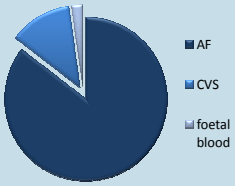
- Multiple congenital abnormalities
- Single abnormality
  - with high risk of chromosome disorder
  - with therapeutic consequences

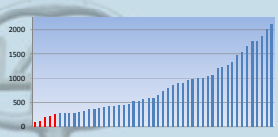
MCA	51
Isolated	41
CDH	45



Sample Details

- Sample size; Mean 7ml AF (2-10ml)
- DNA Yields; Mean 799ng, Median 586ng
- MCC = 2
- Insufficient AF for DNA extraction = 4
- 7 direct AF required culture for result



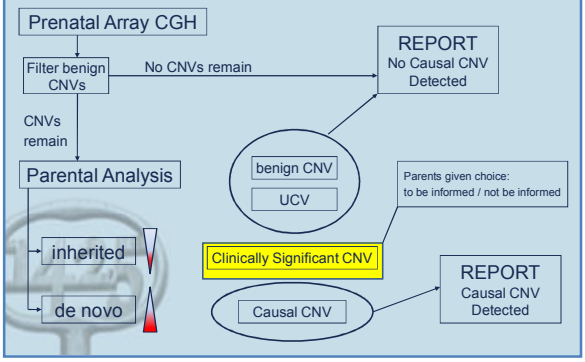


	Direct	Culture
AF	75	42
CVS	11	6
Blood	3	0

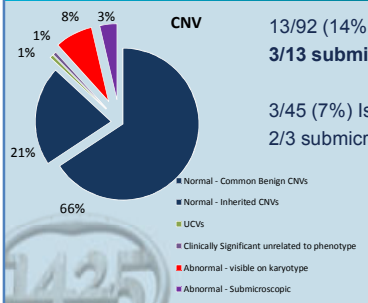
Array Platform

- 4x 180K oligo array (Oxford Gene Technology, UK)
- Targeted at >400 genes & >200 syndromes
- Backbone resolution 150kb
- Targeted resolution 10kb (apply 5kb filter)
- Currently in routine use for postnatal referrals

Workflow for Interpretation & Reporting



Prenatal Results Summary

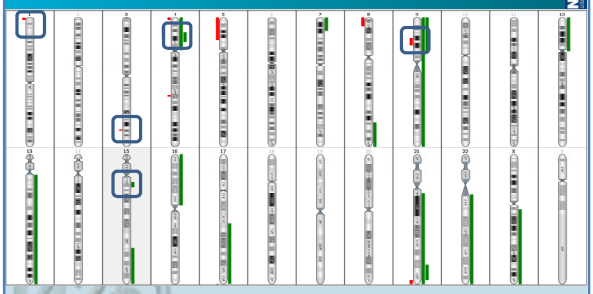


13/92 (14%) Leuven referral cohort
3/13 submicroscopic (23%)

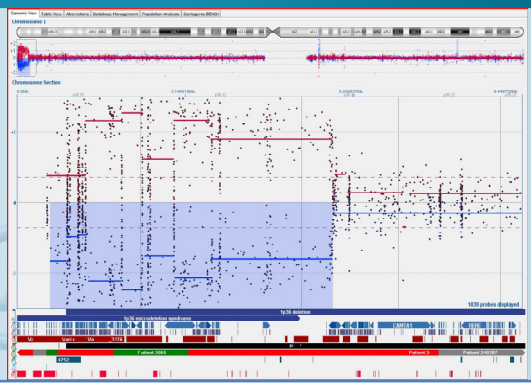
3/45 (7%) Isolated CDH cohort
2/3 submicroscopic

- Karyotyping did not detect any imbalances which were not detected by array CGH
- Additional information from array in 5 cases with abnormal karyotype

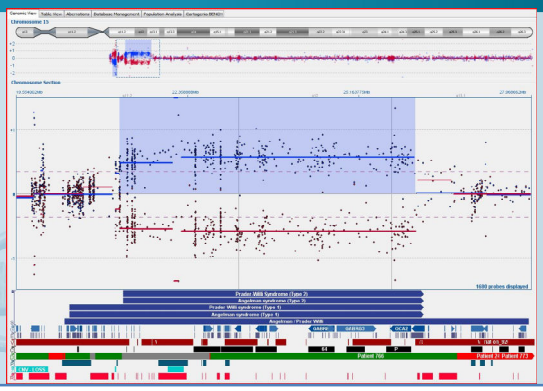
Ideogram of Causal Imbalances



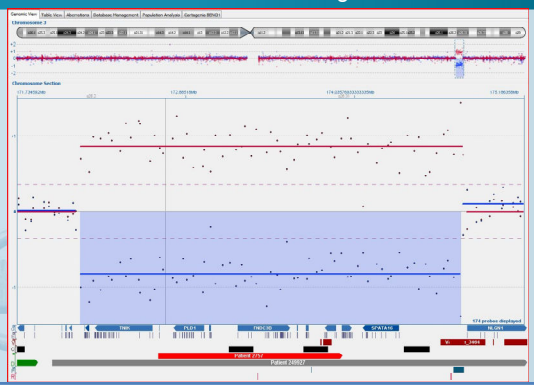
Submicroscopic 1p36 deletion – known syndrome



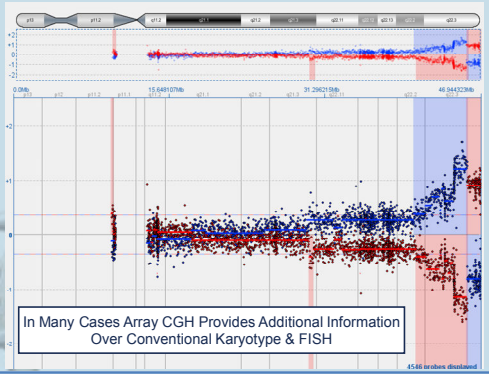
Submicroscopic 15q11.2 duplication (PWAS region)



Submicroscopic 3q26 deletion Non-Recurrent Region



Complex intrachromosomal rearrangement of chr 21



Are We Ready For Prenatal Arrays in Routine Diagnostics?

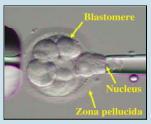
- Array CGH detects causal submicroscopic imbalances
- Prenatal array CGH is ready for routine use*
- Arrays can replace conventional karyotyping*

YES

- With appropriate pre & post test counselling
- For MCA and for severe isolated abnormalities
- UCVs should not be reported (during pregnancy)
- Parents will be given the option to be informed of incidental findings of clinical significance
- Array CGH will be complemented by FISH analysis for positional information, and karyotyping where appropriate
- Consensus guidelines for the use of prenatal arrays are required as this technology enters diagnostic use

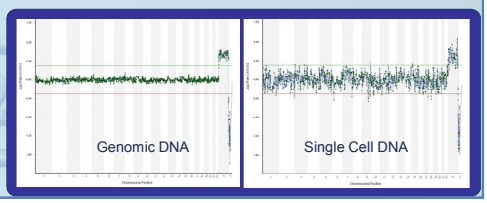
From Prenatal to Pre-Implantation Molecular Karyotyping

Towards pre-implantation genetic diagnosis?

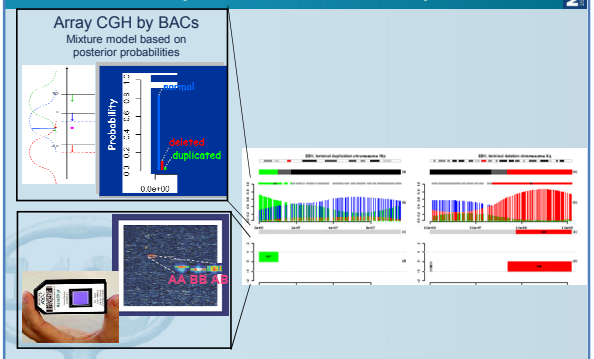


Single blastomere
7 pg DNA

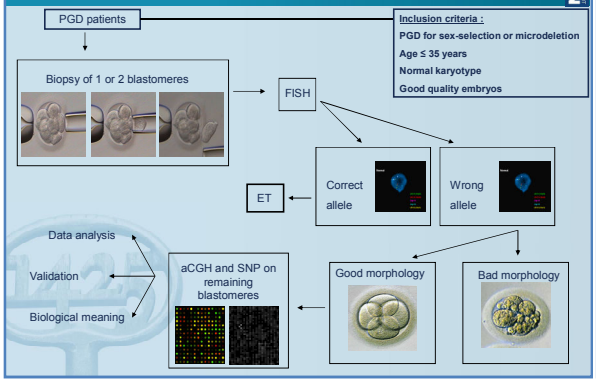
WGA - Genomi Phi
2.5 µg DNA



New methodology: Combine array CGH and SNP array data

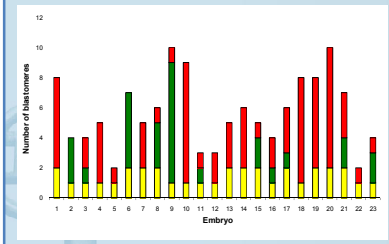


Analysis of human embryos : study design



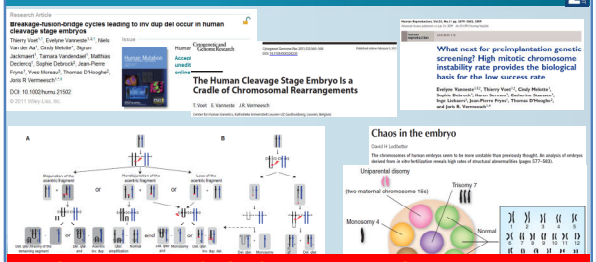
The majority of human cleavage stage embryos contain chromosomally imbalanced blastomeres

- 2/23 (9%) : normal diploid in all cells
- 1/23 (4%) : diploid, but UPID
- 8/23 (35%) : mosaic diploid/aneuploid (4 embryos : ratio diploid/aneuploid > 1
- 12/23 (52%) : mosaic aneuploid (3 embryos : meiotic (same aberration in all cells)



Array = Abnormal
Array = Normal Diploid
FISH (Normal)

Embryos Are Chromosomally Unstable



The Clinical Utility of PGD for Aneuploidy Screening in Cleavage Stage Embryos is Questionable;
- Mitotic error rate is higher than meiotic error rate
- One cell is not representative of whole embryo

ACKNOWLEDGEMENTS

CENTRUM MENSELIJKE ERFELIJKHEID
UZ LEUVEN

Centre for Human Genetics
KU Leuven / UZ Leuven

Laboratory for Cytogenetics & Genome Research
Prof. Dr J. Vermeesch (H.o.D.)
C. Melotte

Prenatal Section
Dr K. Van den Bogaert
E. Mattheeuws

Clinical Geneticists
Prof. Dr J.P. Fryns (H.o.D) – CME)
Prof. Dr K. Devriendt
Prof. Dr Th. de Ravel de l'Argentière
Prof. Dr H. Van Esch
Prof. Dr E. Legius

Referring Clinicians

Patients and Families

Leuven University Fertility Centre

Fetal Medicine Team
Prof. Dr J. Deprest (H.o.D.)
Dr P. DeKoninck
Prof. Dr L. de Catte
Prof. Dr I. Witters

Vermeesch Research Group
Ass. Prof. Dr T. Voet
Dr E. Vanneste
Dr F. Hannes
Dr J. Van Houdt
Dr B. Nowakowska
N. Van der Aa
M. Zamani Esteki
S. Jackmaert
L. Bittman
C. Robberecht
G. Peeters

ESAT – SCD:SISTA/COSIC/DOCARCH
Prof. Dr Y. Moreau (H.o.D.)

Thank You for your attention

WORKSHOP TRAINING SESSIONS

2-D DIGE training session

I. 2-D DIGE experimental design
II. Practical aspect of performing DIGE experiment
III. DIGE data analysis

2-D Fluorescence Difference Gel Electrophoresis (2-D DIGE) is a method that labels protein samples prior to 2-D electrophoresis, enabling accurate analysis of differences in protein abundance between samples. 2-D analysis experiments commonly address questions like protein level differences caused by a disease state, drug treatment, life-cycle stage etc. Some protein level differences are small and the results are affected by experimental variation originating both from the system and from inherent biological variation.

System related result variation may arise either from gel-to-gel variation, which can result from differences in electrophoretic conditions between first dimension strips or second dimension gels, gel distortions, sample application variation and user-to-user variation, of from variation due to user-specific editing and interpretation when using the data analysis software.

Inherent biological variation arises from intrinsic differences that occur within a population. For example, differences from animal-to-animal, plant-to-plant or culture-to-culture which have been subjected to identical conditions.

System variation cannot be overcome when using conventional 2D electrophoresis but by using DIGE system it is possible to minimize the gel to- gel variation effects on results by using a combination of multiplexing of samples and inclusion of an internal standard within each gel.

Biological variation cannot be removed from any 2D electrophoresis experiments. However, using DIGE system allows the inherent biological variation to be effectively differentiated from induced biological changes using highly accurate measurement of protein abundance changes, an appropriate experimental design and advanced statistical analysis. The use of biological replicates in the experimental design ensures a true measurement of induced biological differences above the background of inherent biological variation.

When designing 2D DIGE experiments, the following recommendations should be considered:

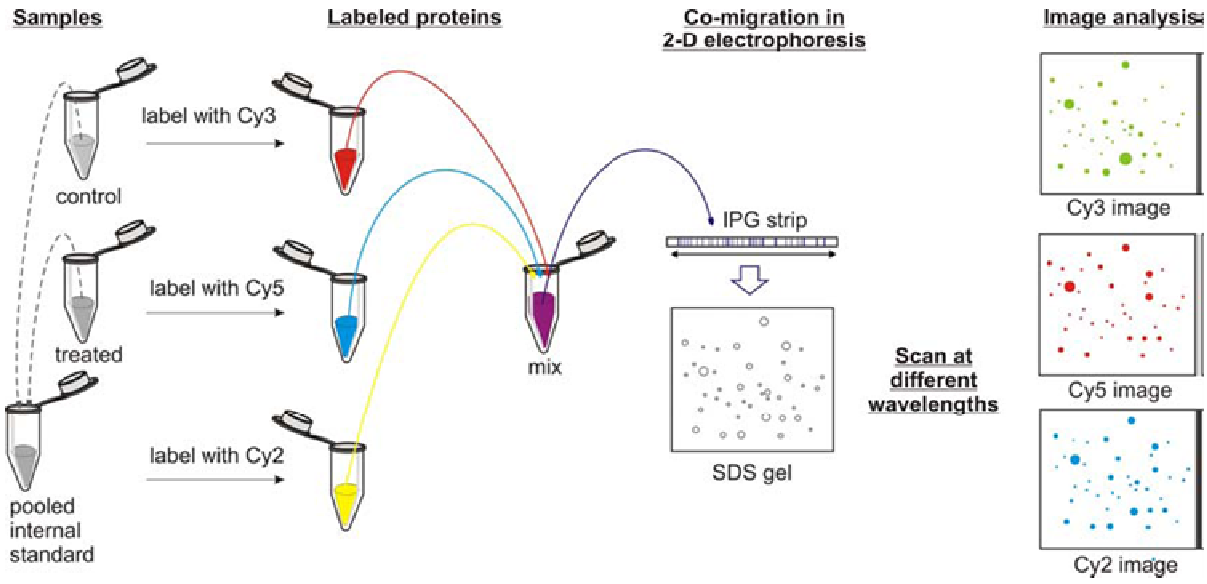
1 Inclusion of an internal standard sample on each gel

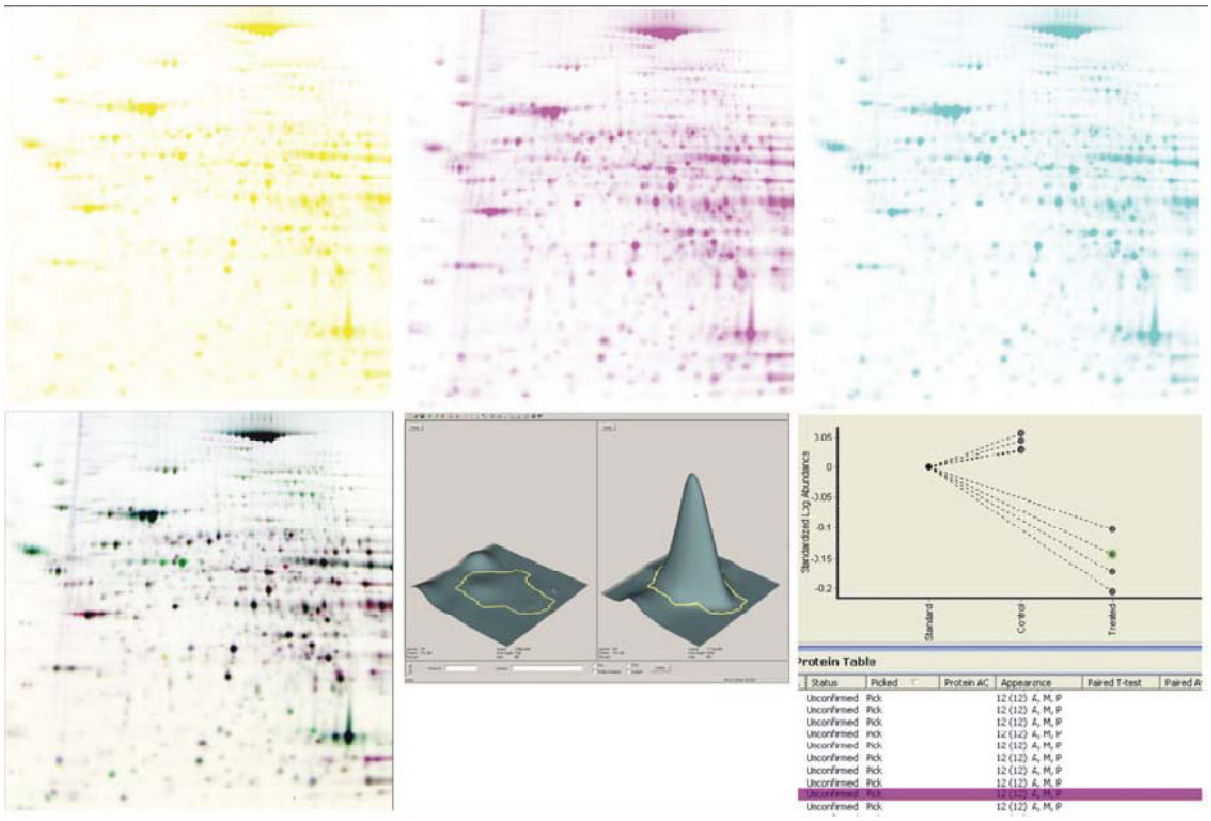
2 The requirement for biological replicates such as multiple cultures, tissue etc.

3 Randomization of samples to produce unbiased results

4 No gel replicates of the same sample is needed

From Sample Labeling to Scanning





Planning of the Experiments

Perform a "planned randomization" of Dye labeling and application on IPG strip.

	Int. stand. / Cy2	Sample / Cy3	Conc (µL of 50 µg)	Sample / Cy5	Conc (µL of 50 µg)	IPG strip no.	Grad. cm
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							

Sample Preparation

→ Add DIGE "Lysis"-Buffer (labeling buffer) to each sample to a concentration in the range **5-10 mg/mL**

Independently from the experiment size, at least **75 µg** of each sample is required: **50 µg** for sample labeling, **25 µg** for creating the internal standard.
If accurate quantification of the samples´ protein amount could not be determined: under labeling is not an issue. Only over labeling would create multiple labels in the basic area.

Quantification

It is recommended to use the ***Bradford*** method

Clean up

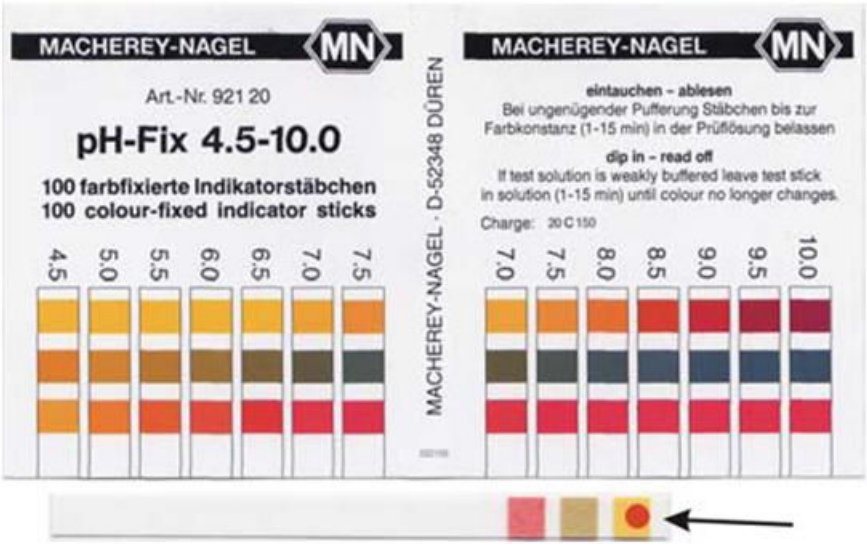
It is recommended to use the ***Ettan™ 2-D Clean-Up Kit***

pH control and adjustment

For efficient labeling the protein sample *should* have optimally pH 8.5; it *must* be above pH 8.0.

→ Check the pH carefully: pipette **2 µL** sample on a pH indicator paper. Read out the pH value immediately, because the color will shift with time.

If necessary, adjust the pH value with adding **100 mM** NaOH solution.



Note: *Proteins have some inherent buffering capacity and may have decreased the pH value of the sample solution below pH 8. Samples which have been cleaned up with TCA acetone or the Ettan™ 2-D cleanup kit can be acidic.* ***Beware:*** *Sometimes, when samples have been transported in dry ice and the tubes have not been sealed well enough, CO_2 has diffused into the samples and caused a strong drop of the pH value. In this case it might be necessary to adjust the pH value with 250 mM NaOH.*

CyDye Labeling

Reconstitution of the CyDyes

Use 99.8% anhydrous Dimethylformamide (DMF) less than 3 months old from day of opening. The quality of the DMF is critical to ensure that the protein labeling is successful. The DMF must be anhydrous and every effort should be used to ensure it is not contaminated with water.

Stock dye solutions

Reconstitute CyDye minimal dyes solid compounds DMF to a concentration of:

1 nmol/µL

e.g. 5 µL DMF to 5 nmol/L of dye.

The solutions are stable at –20 °C for several months.

→ Take a small volume of DMF from its original container and dispense into a microcentrifuge tube.
→ Take the CyDye from the –20 ºC freezer and leave to warm for 5 minutes at room temperature.
→ After 5 minutes: add **5 µL** of the DMF to each new vial of CyDye.
→ Replace the cap on the dye microcentrifuge tube and vortex vigorously for 30 seconds.

→ Centrifuge the microcentrifuge tube for 30 seconds at 12,000 g in a benchtop microcentrifuge.

The dye can now be used.

Working dye solutions

→ Briefly spin down dye stock solution in a microcentrifuge
→ Dilute 1 volume of the stock CyDye in 1.5 x volumes of high grade DMF to create:

400 pmol/µL

→ Add **3 µL** of the DMF first to the sterile microcentrifuge tube.
→ Add **2 µL** of the stock dye and mix.

Now you have 2,000 pmoles CyDye in 5 µL.

Labeling of the samples and the internal standard

Samples:

→ Add a volume of sample equivalent to **50 µg** protein to a microcentrifuge tube.
→ Add **1 µL** of diluted CyDye to the microcentrifuge tube containing the sample.
→ Mix and centrifuge briefly in a microcentrifuge. Leave on ice for **30 minutes** in the dark.
→ Add **1 µL** of 10 mmol/L lysine to stop the reaction. Mix and spin briefly in a microcentrifuge. Leave for **10 minutes** on ice in the dark.

Internal standard

n is the *number of gels* in the experiment

→ Add a volume of pooled internal standard equivalent to ***n*** × **50 µg** protein to a microcentrifuge tube.
→ Add ***n*** µL of diluted Cy2 to the microcentrifuge tube containing the pooled standard (i.e. 300 µg of protein would be labeled with 2,400 pmoles of dye).
→ Mix and centrifuge briefly in a microcentrifuge. Leave on ice for 30 minutes in the dark.
→ Add ***n*** **µL** of 10 mmol/L lysine to stop the reaction. Mix and spin briefly in a microcentrifuge. Leave for 10 minutes on ice in the dark.

Labeling is now finished:

400 pmol per 50 µg protein

Samples and pooled standard can now be stored for at least three months at −70°C in the dark.

Preparation for loading the samples onto the IPG strips

→ Combine the labeled samples and pooled internal standards according to the experimental design.

→ **Add an equal volume of 2 x lysis buffer** (containing the IPG buffer and the DTT) to each sample and standard, and leave on ice for 10 minutes.

→ *If necessary*, dilute samples further with a 1:1 mix of DIGE "Lysis"-Buffer (labeling buffer) and 2 x lysis buffer to *a minimum of 100 µL* for optimum protein entry.

In highly concentrated samples some proteins tend to aggregate and precipitate.

Isoelectric Focusing

Set up running conditions

→ Enter the running conditions in the computer.

→ Start Ettan™ IPGphor3 program.

→ Select the instrument connected (usually instrument 1 of four)

→ Select pI range, strip length and number of strips. A programmed voltage running curve will show up.

→ Click on the "table" icon on the right low position of the screen (red arrow).

→ Save method under a new name.

→ Transfer to IPGphor instrument.

Preparing second dimension gels

→ During unloading rinse the cassettes with tap water to remove excess polyacrylamide.
→ Inspect each gel cassette for eventual air bubbles. Gels with air bubbles should not be used.
→ Place each cassette into the cassette rack.

Preparing the equilibration solution

The frozen solution aliquoted in 50 mL centrifuge tubes need some time to thaw.

→ Take **200 mL** of equilibration buffer from the freezer and let them thaw at room temperature.
→ Weigh out **1 g** DTT and **2.5 g** iodoacetamide.

Refocus the IPGstrips

→ Refocus the proteins by applying 10,000 V on the strips for 15 minutes before equilibration *(only necessary when strips left in IPGphor for longer than ½ hour).*

Setting up the Ettan™ DALT*six* modular system

→ Take the cassette carrier and the upper buffer chamber out of the instrument.
→ Connect the tubing to a circulating thermostat, which has been set to 25 °C.
→ Pour **450 mL** running buffer (10 x conc.) into the lower buffer tank.
→ Fill **4 L** MilliQ water into the tank.
→ Plug the cable of the pump in. The pump starts to circulate the liquid, mixes the concentrate with the water.

Equilibration of the IPG strips

→ Remove the electrodes, the loading cups, and the electrode pads from the Manifold.
→ Pour the dry strip cover fluid out from the Manifold.
→ Add **1 g** DTT to **100 mL** equilibration buffer, mix thoroughly and pour into the Manifold.
→ Place the manifold on an orbital shaker for 15 minutes, which is set to 30 rpm.
→ After 15 min, pour out the first equilibration buffer.
→ Add **2.5 g** iodoacetamide to **100 mL** equilibration buffer, mix thoroughly and pour into the Manifold.
→ Place the Manifold on an orbital shaker for another 15 minutes, which is set to 30 rpm.
→ After 15 min, pour out the second equilibration buffer.

Do not leave the strips longer in equilibration buffer, because this would elute a part of the proteins from the strip.

Application of the IPG strips onto the SDS gels

→ Heat pre-prepared agarose on a heating stirrer or in microwave oven to melt it.
→ Pour a few milliliters of MilliQ water on the upper gel edge using a squeeze bottle. This will greatly facilitate the insertion of the IPG strip into the cassette.
→ Lay the cassette on the bench with the longer glass plate down, the protruding edge oriented towards the operator.
→ Place the IPG strip with the acidic end to the left, gel surface up onto the protruding edge of the longer glass plate as shown in the figure below.

The correct orientation of the IPG strip is particularly important for Spot-picking gels (one glass plate treated with Bind-silane)

Inserting the cassettes

→ Wet the buffer seal of the upper buffer chamber with 0.1 % SDS water. Spray with the plant sprayer used for overlaying the gel edges.
→ Insert the cassettes into the cassette carrier and place it into the tank.

When less than 6 gels are run, insert blank cassettes into the free positions. In the front, however, a gel cassette should be inserted: this makes it easier to watch the migration of the Bromophenol Blue front during the run.

When all cassettes are in place, the level of the anodal buffer should have reached the mark "LBC start fill". If necessary, add MilliQ water to reach the mark. The filling procedure is shown in detail in the figure below.

→ Fill the upper buffer chamber with the **1.4 L** 1x running buffer.
→ Immediately fill **5.6 L** of 1 x concentrated running buffer into the lower buffer tank to the same level in order to establish a hydrostatic balance.

Running conditions

	Step 1	10 mA per gel	≈0.5 W/gel	1 hour
Fast run	Step 2		15 W/gel	ca. 4 hours
Over night run	Step 2	25 mA per gel	1.5 W / gel	Over night

Scanning with Ettan DIGE Imager

→ Turn on the scanner and the computer.
→ When **READY** light is on, double click the **Ettan DIGE icon** on the desktop to start the DIGE software.
→ Clean the gel glass plates with distilled water using a lint-free tissue. It is important that the glass plates are clean, dry and free from lint.
→ Insert the dried cassette into the EDI cassette with the small plate facing down, resting in the seal at the bottom of the cassette. The large plate should hang over the area at the rear of the cassette.

Note: The acidic side of the IPGstrip should point to the left. If the acidic side points to the right, the image needs to be flipped later on: either in the EDI *Report Viewer* or later with *ImageQuant*.

→ Put the lid on the cassette, and close the cassette by turning the locking cams up.
→ Slide the Scanner door until it is fully open.
→ Insert the cassette into the scanner with the cams towards you and tilted up, so that the cassette fits under the beveled edge of the carrier. Push it down and in on the front of the cassette to lock it in place. Close the Scanner door.
→ Start the Scanner software by double clicking the **Ettan™ DIGE Imager icon** on the computer desktop.

Select scan parameters

→ Select the appropriate gel format.
→ Specify the type of chemistry used for your sample by selecting an item in the *Chemistry list*.
→ Select the pixels size, i.e. resolution to use, in the *Pixel size list*. 100 µm are selected for analysis by DeCyder™.
→ Select the number of channels to be scanned by clicking the *Channel check box*.
→ Select DIGE file naming format, and have a standard defined in your experiment, specify which dye (usually Cy2) represents the standard under Standard.
→ Select the exposure time to be used for each channel in the *Exposure time list*.
→ Perform a quick test scan on a small area to identify a suitable exposure time for each channel. Therefore select an area containing the most intense spots. Saturated areas are shown in **red**.

Scanning

→ Click the *Scan button* to start the scan.
→ In the *File Name Dataset* dialog box, browse to a directory and enter a name for the dataset, or choose New folder to create a new folder.
→ Click *Proceed* to start the scan.

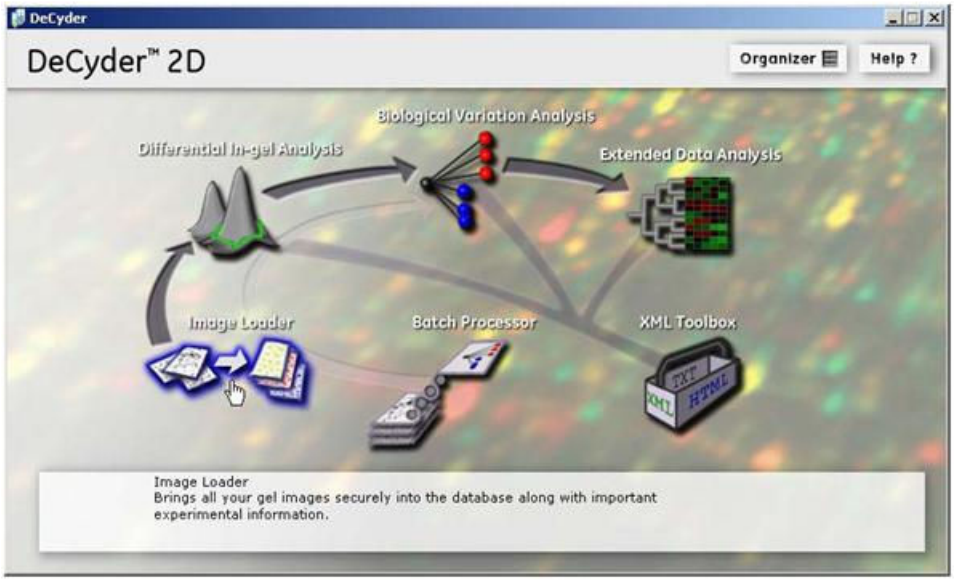
Note: When scanning more than one channel, scan data files (in a .gel format) are created in the dataset directory. Two identical index files (in a .ds format) are also created. One index file resides in the dataset directory. The other index file resides at the same level as the dataset directory. Scanning of one channel only creates data files in a ******.**gel** format.

DeCyder™ 2D Analysis

Image loading

DeCyder™ sits on an Oracle Database, images must first be uploaded before they can be analysed.

→ Start the Image Loader from the DeCyder main screen.



Click Add button – browse for images.
Select images for loading, highlight all images to be loaded and press open.

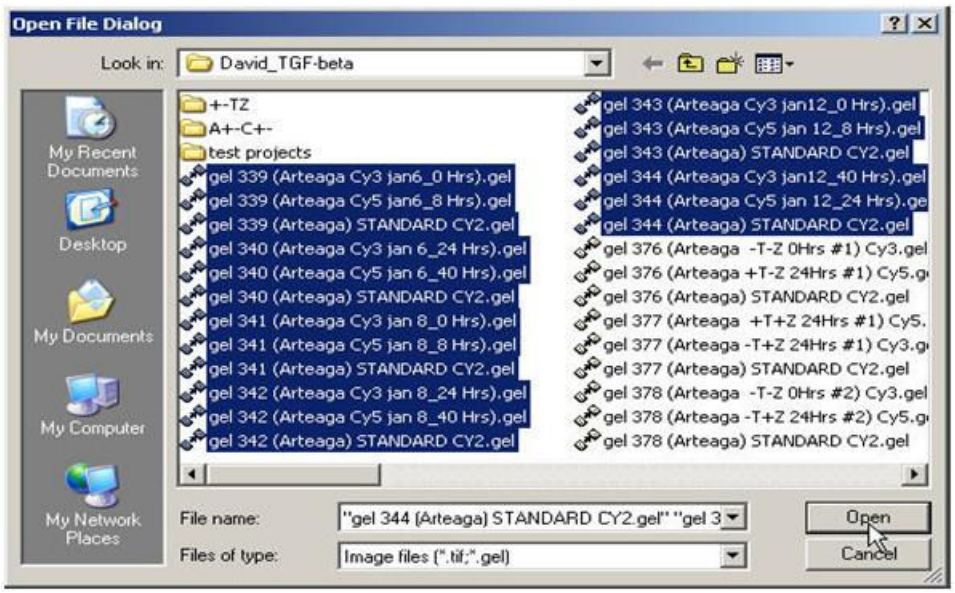


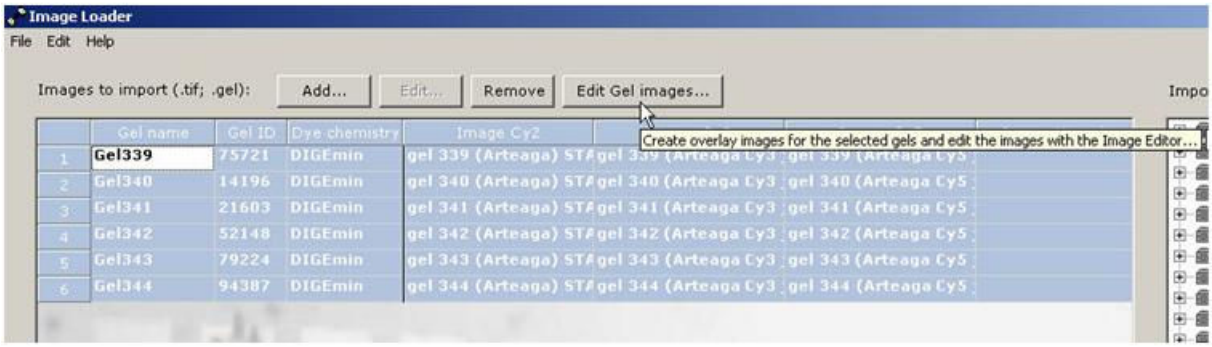
Image Cropping

- All areas that do not contain useful information should be removed.
- Ensure that all relevant spots remain inside the image.
- It is more important to ensure **consistent patterns** than equal sizes.

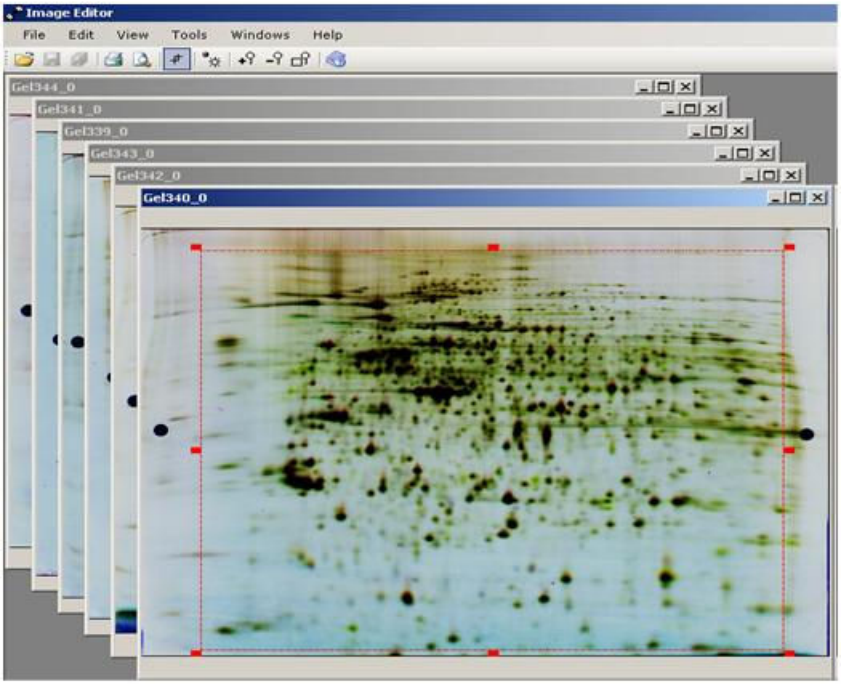
If your images are already correctly cropped go on with Differential In-Gel Analayis after image loading.

If your images contain areas like vertical streaks at the pH extremes, buffer front etc., use the Image Editor from the Image Loader to perform correct cropping.

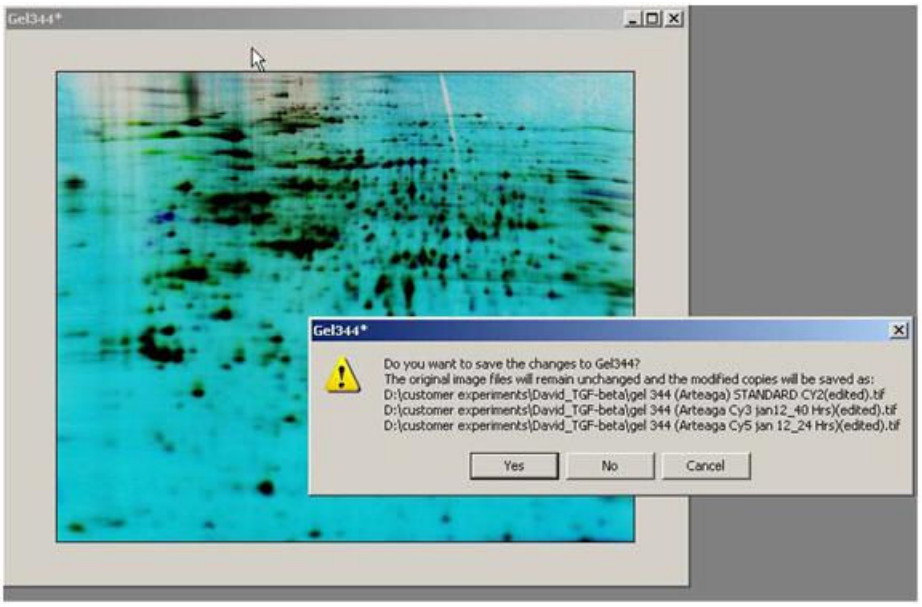
→ After images are added within the Image Loader select all images and press Edit Gel images.



→ All images will be displayed as false colour overlay in a stack mode.
→ Select the first gel by clicking on the gel bar and use the cropping tool to define the crop region.



→ After the cropping area is defined double-click the left mouse button inside the frame to perform cropping.



The cropped images will be saved with the add-on '(edited)'. The next gel in order will start with the previous cropping frame, which can be edited before cropping is performed. Until the last gel is cropped import them into a selected or newly created project.

→ Close Image loader

Differential In-Gel Analysis (DIA)

The DIA (**D**ifferential **I**n-Gel **A**nalysis) is the **Initial step in DeCyder analysis**:

- performs spot co-detection

- spot quantification by normalization and ratio calculation

This module of DeCyder™ removes most of the system variations

Viewing – contrast setting

This measure makes you aware of any disturbances which might effect spot detection and matching.

low contrast high contrast

Process gels

Consists of:

- spot co-detection
- normalization (see Appendix C in DeCyder user manual)
- ratio calculation

Spot detection

→ Select V5 or V6 spot detection algorithm (V6 is default).

→ Estimated number of spots.

→ Select **10,000** (overestimation) to compensate for the detection of non-protein objects on the image, e.g. dust particles

which are subsequently excluded from the analysis.

The spots are now detected…..

Exclude filter

→ Select **30,000** for volume (very good exclusion filter for normal quality gels, might be adjusted for different quality)

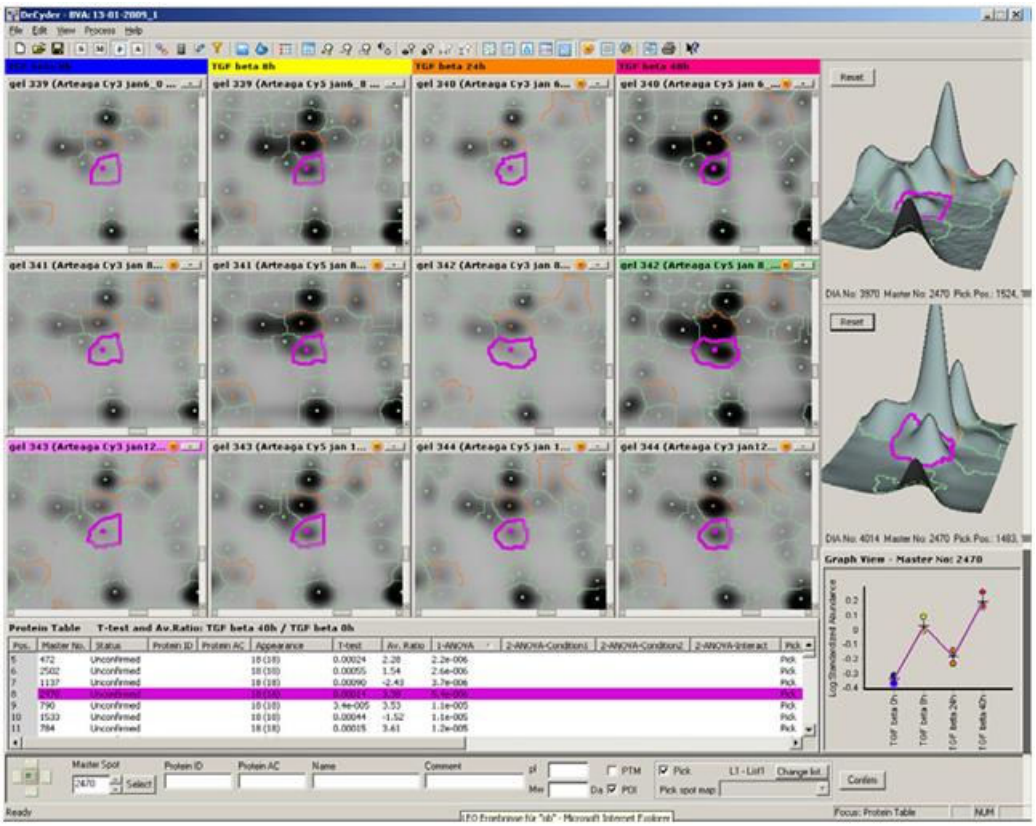
Do not apply the other filters. For normal quality images other filters are not needed.

After exclusion of non-protein objects the DIA module is finished. This analysis could be done with all gels one by one, but it is preferably performed with the batch processor.

B*iological* V*ariation* A*nalysis (BVA)*

The **B**iological **V**ariation **A**nalysis (BVA)

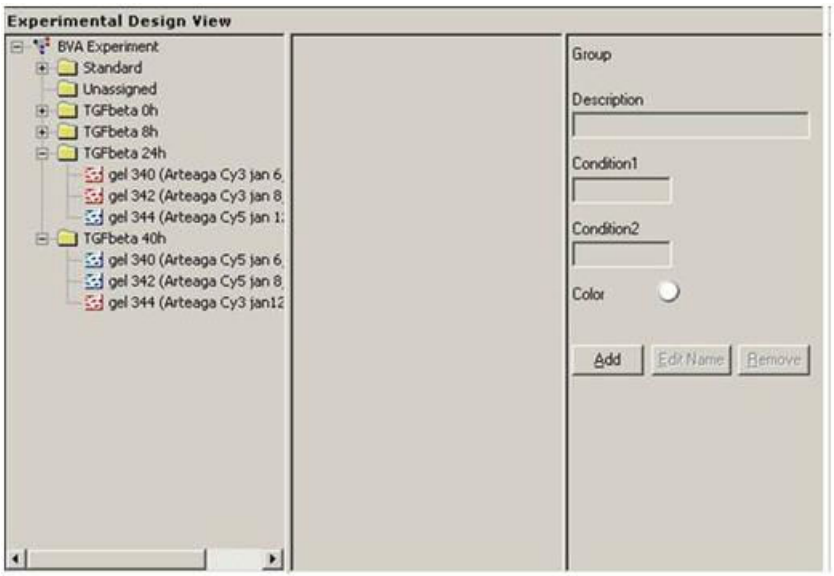
- processes multiple gel images
- performs gel to gel matching of spots
- allowing quantitative comparisons of protein expression across multiple gels.

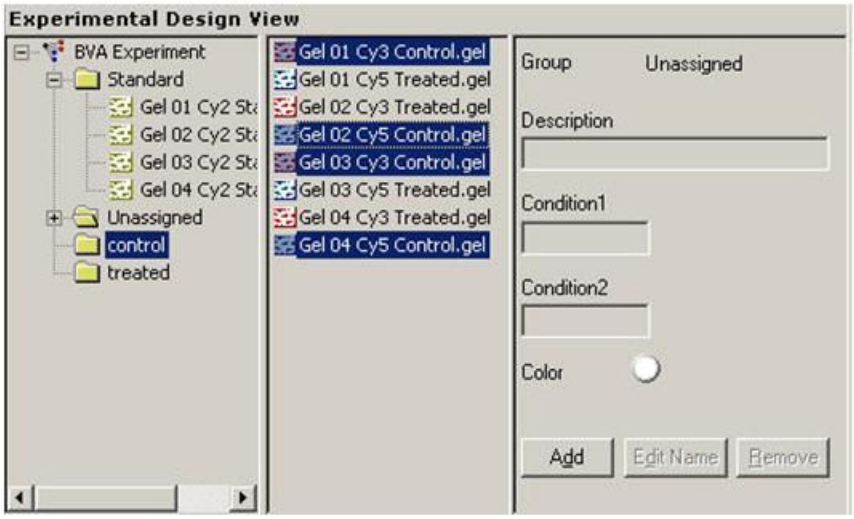


Create workspace/Import DIA-workspace

→ Select DIA workspaces in browser, highlight , create…

Assign experimental groups (in S mode)

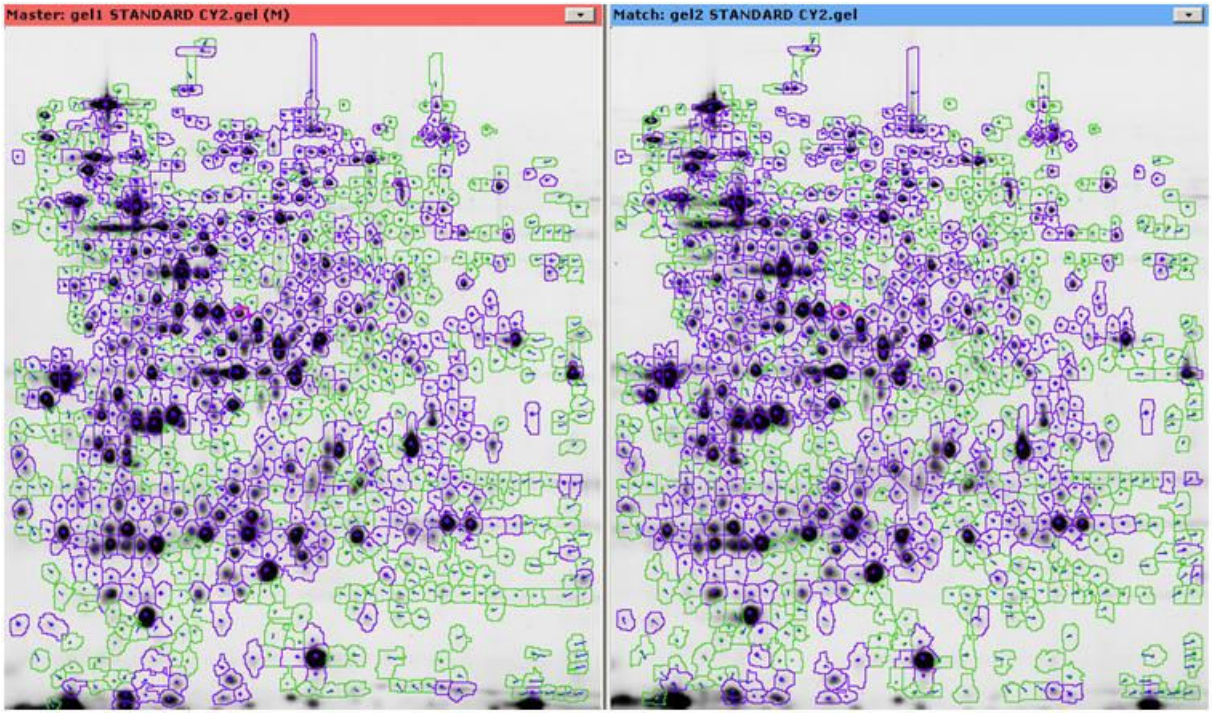




Define **master** for matching:

→ Check standard images for quality: select the most representative image (average quality, not the best, not the worst) and assign it as **master.**

Match gels (in M mode)



Statistical analysis (in P mode)

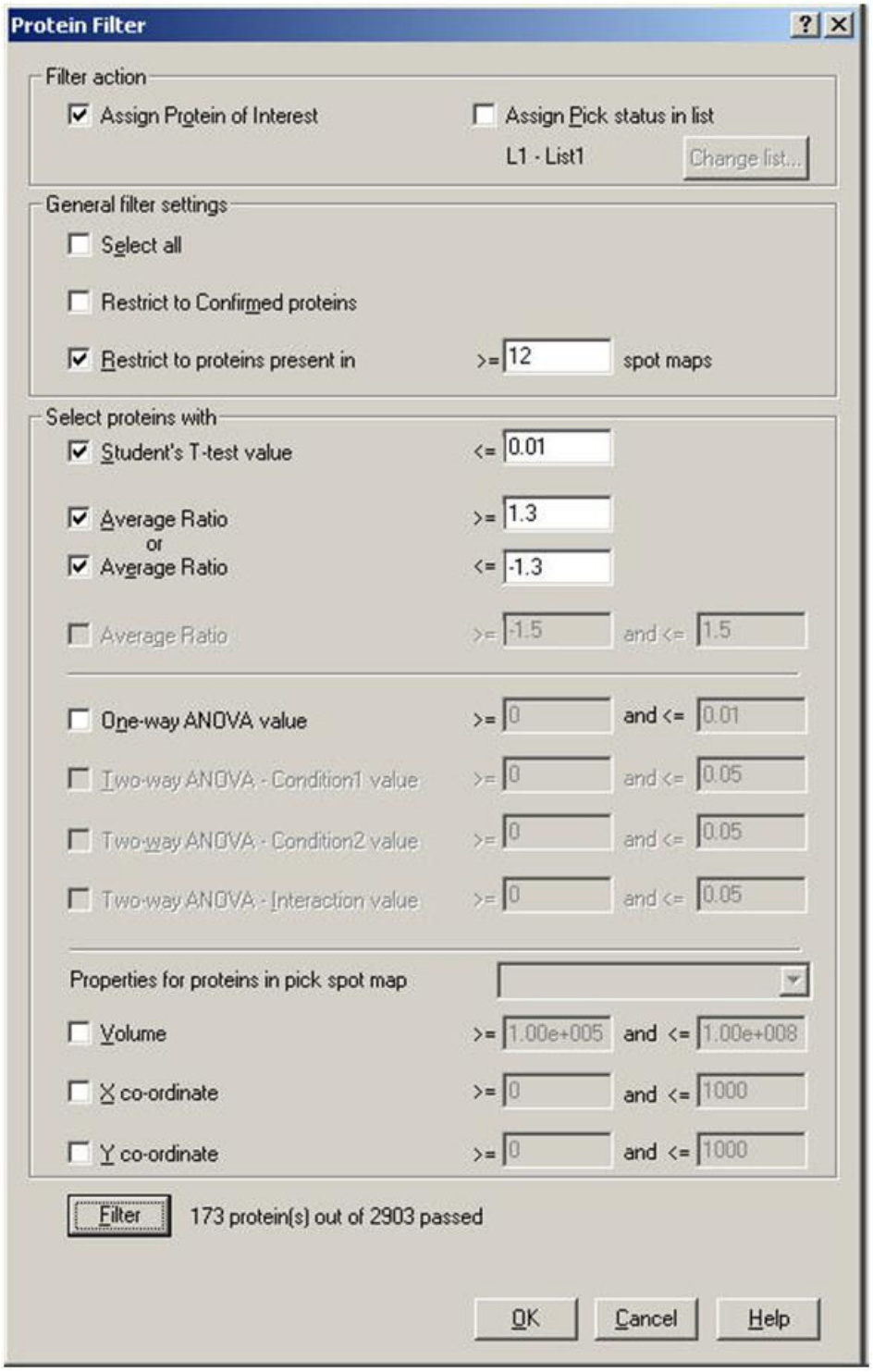
In protein statistics define to calculate average ratio and Student´s T-test for two experimental groups. If more than two experimental groups are analyzed, select additionally One-Way ANOVA.

Typical example of a differentially expressed protein with a clear time dependency.

Use Protein filter to assign *protein of interest* and/or *pick* status (in P mode)

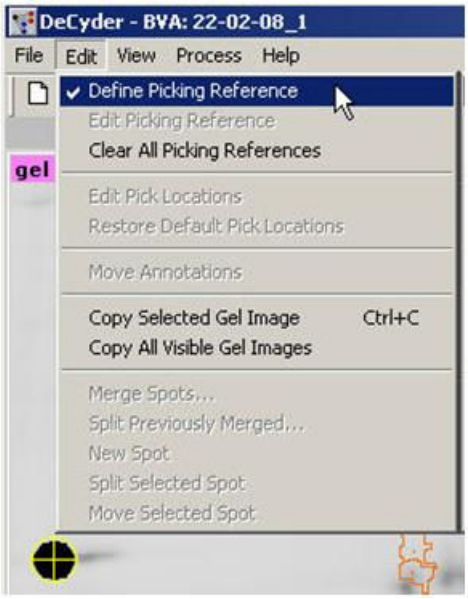
Recommended filter settings:

To filter for differentially expressed proteins between two selected experimental groups. In this example the filter is restricted to proteins which are present in at least 12 out of 18 spot maps (4 of 6 gels).

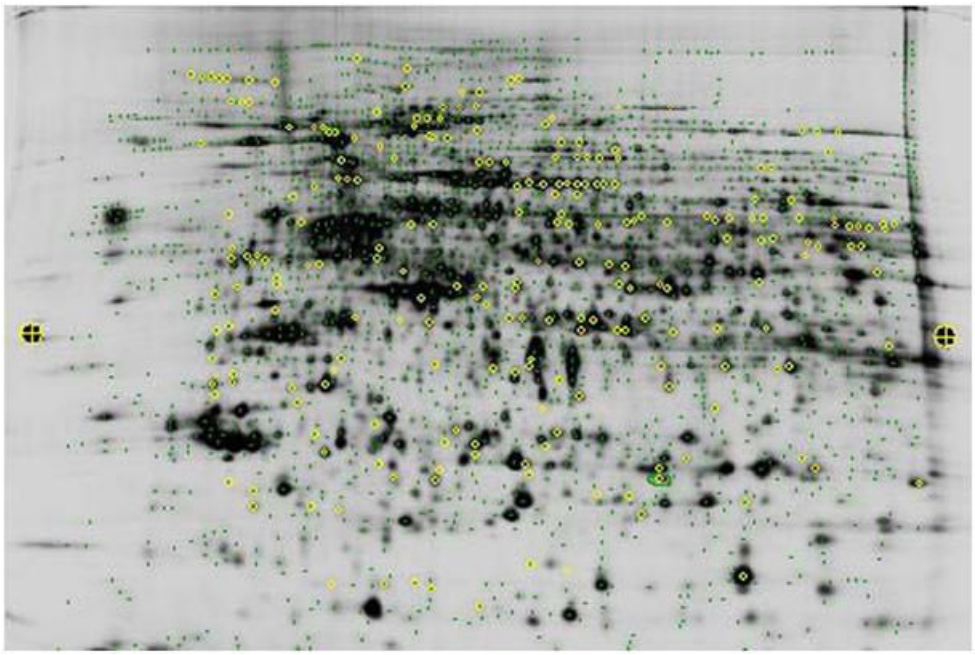

Protein Filter
Filter action
Assign Protein of Interest
Assign Pick status in list
L1 - List1
Change list...
General filter settings
Select all
Restrict to Confirmed proteins
Restrict to proteins present in
>= 12
spot maps
Select proteins with
Student's T-test value
<= 0.01
Average Ratio
>= 1.3
or
Average Ratio
<= -1.3
Average Ratio
>= -1.5
and <= 1.5
One-way ANOVA value
>= 0
and <= 0.01
Two-way ANOVA - Condition1 value
>= 0
and <= 0.05
Two-way ANOVA - Condition2 value
>= 0
and <= 0.05
Two-way ANOVA - Interaction value
>= 0
and <= 0.05
Properties for proteins in pick spot map
Volume
>= 1.00e+005
and <= 1.00e+008
X co-ordinate
>= 0
and <= 1000
Y co-ordinate
>= 0
and <= 1000
Filter
173 protein(s) out of 2903 passed
OK
Cancel
Help

Create Pick List

→ Apply pick function to the pick gel.
→ Define both Picking references.

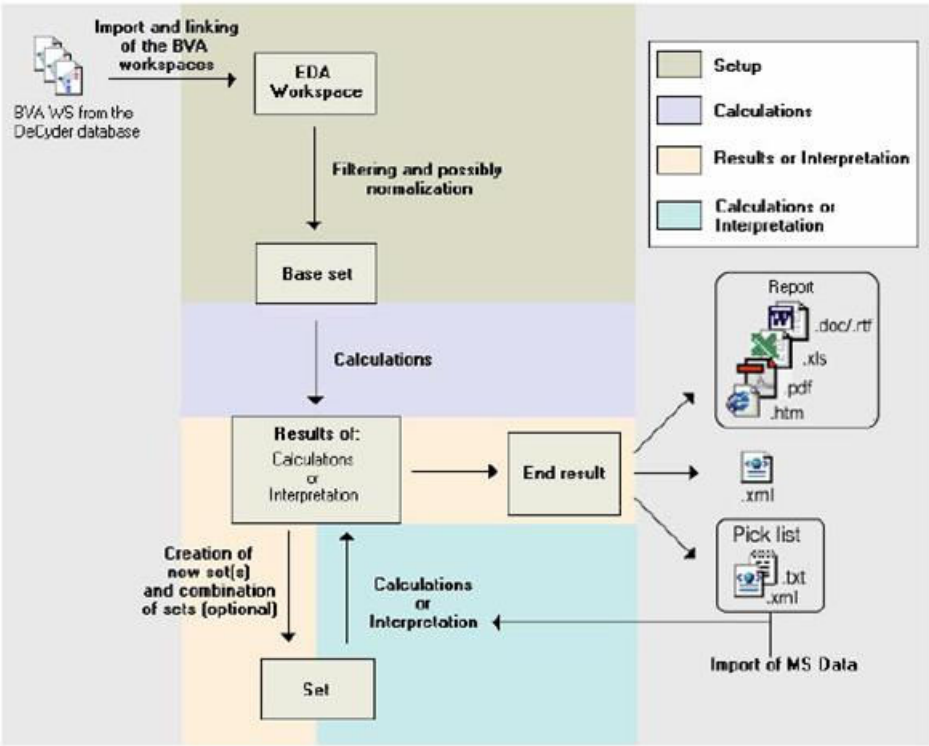


After second picking reference is assigned, all pick positions are indicated.



Extended Data Analysis *(EDA)*

Workflow



→ Select and import one or several BVA workspaces

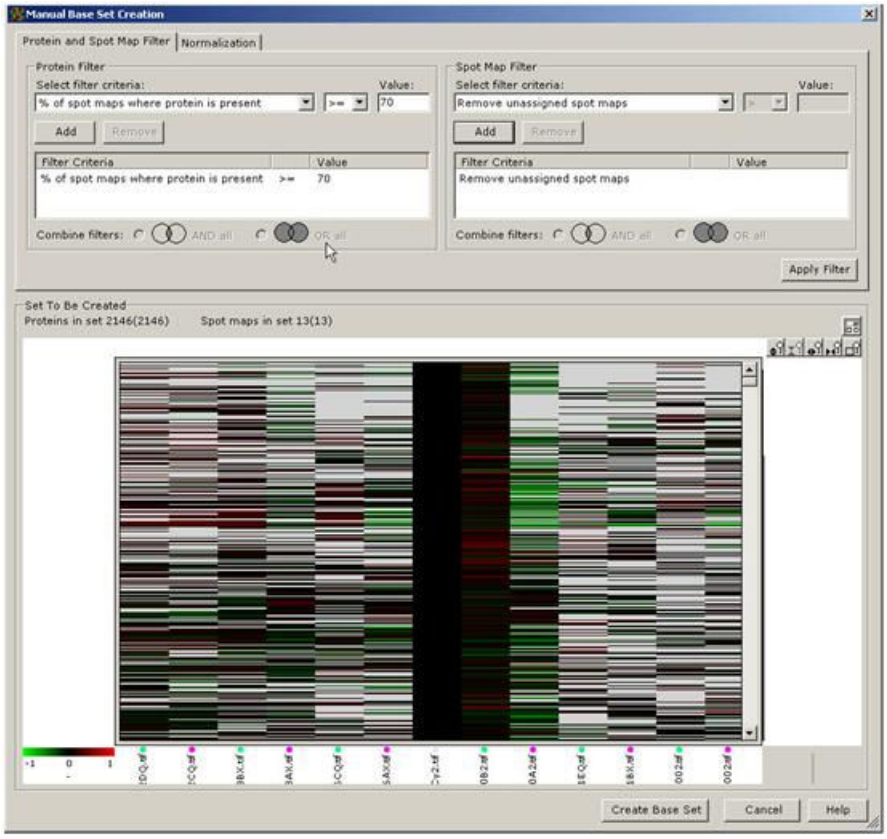
Note: No other values than **standardized log abundance** will be imported.

→ Check experimental design and change if necessary

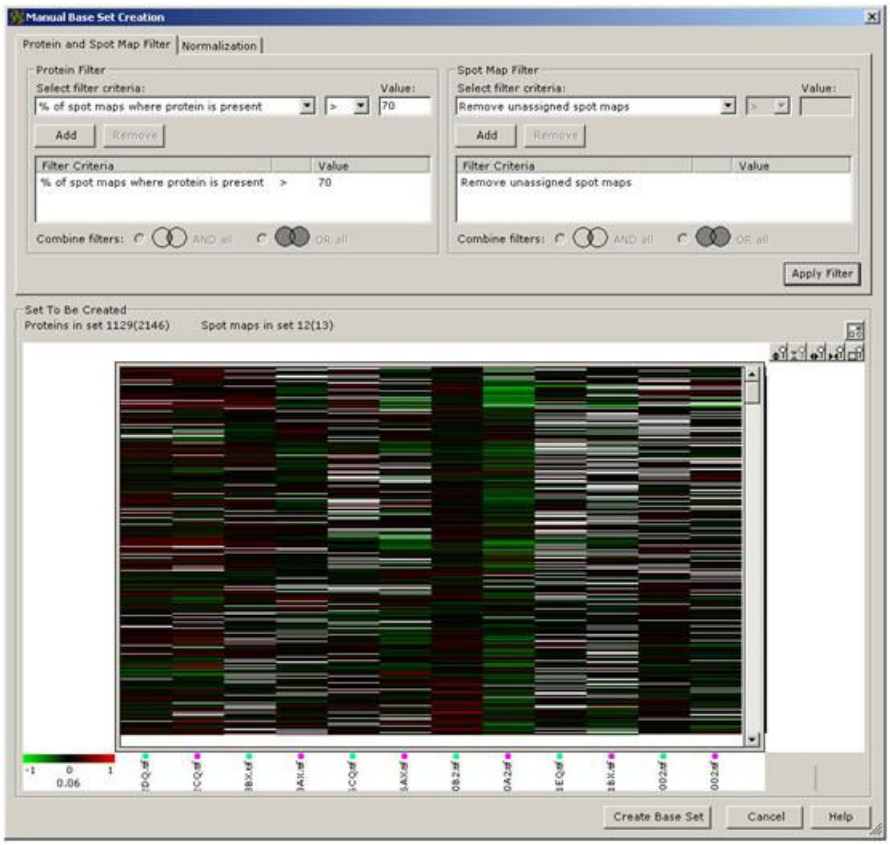
→ Generate "*base set*"

The "*base set*" should consist of proteins which show a minimum frequency of appearance (e.g. in 70 % of the images, *filter 1*). Too many *"missing values (protein spots)"* would cause abort of calculation. Images, which do not belong to an experimental group should be excluded (*filter 2*).

Data set before filtering



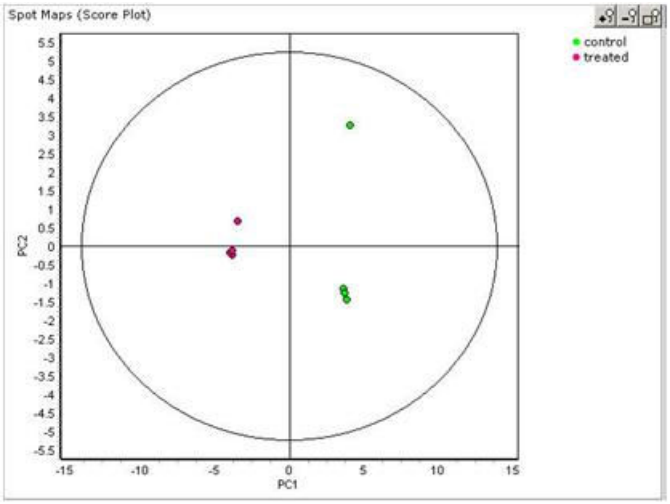
Data set after filtering



Two calculations should be performed:

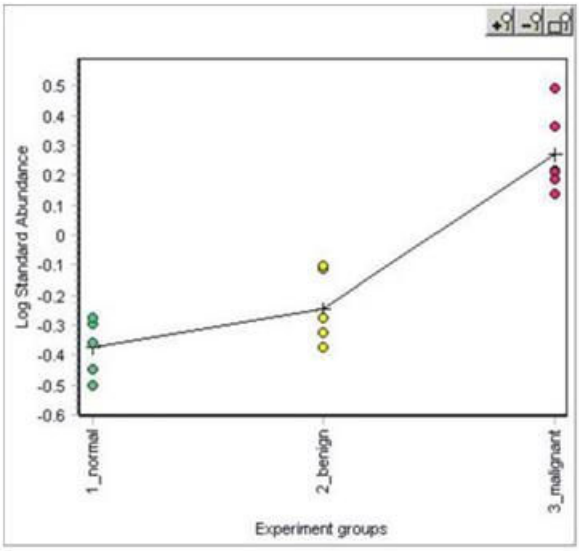
- **Principal Component Analysis** (to get an idea about possible outliers)
- **Differential Expression Analysis** (Student´s t-test, ANOVA) With this we create a tool for the separation of differentially expressed from the non differentially expressed proteins.

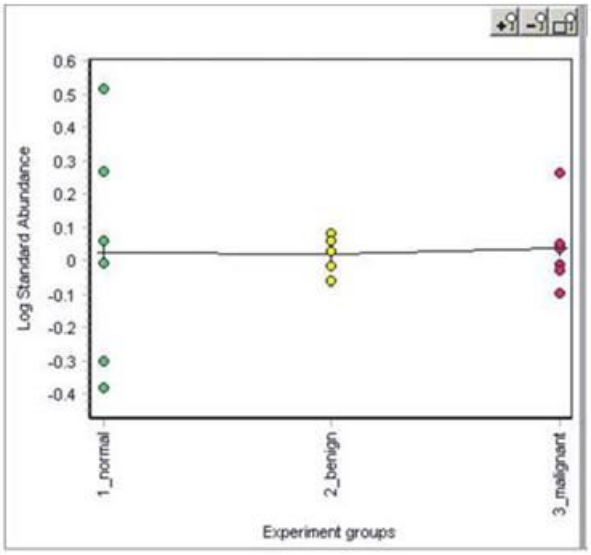
Principal Component Analysis



In this score plot one of the spot maps is separated by PC2 from the other group members. This is obviously an outlier and some investigation should be done to identify the reason for this behaviour. Depending on the questions to answer this spot map might be excluded from further analysis (otherwise it might bias the results).

Differential Expression Analysis



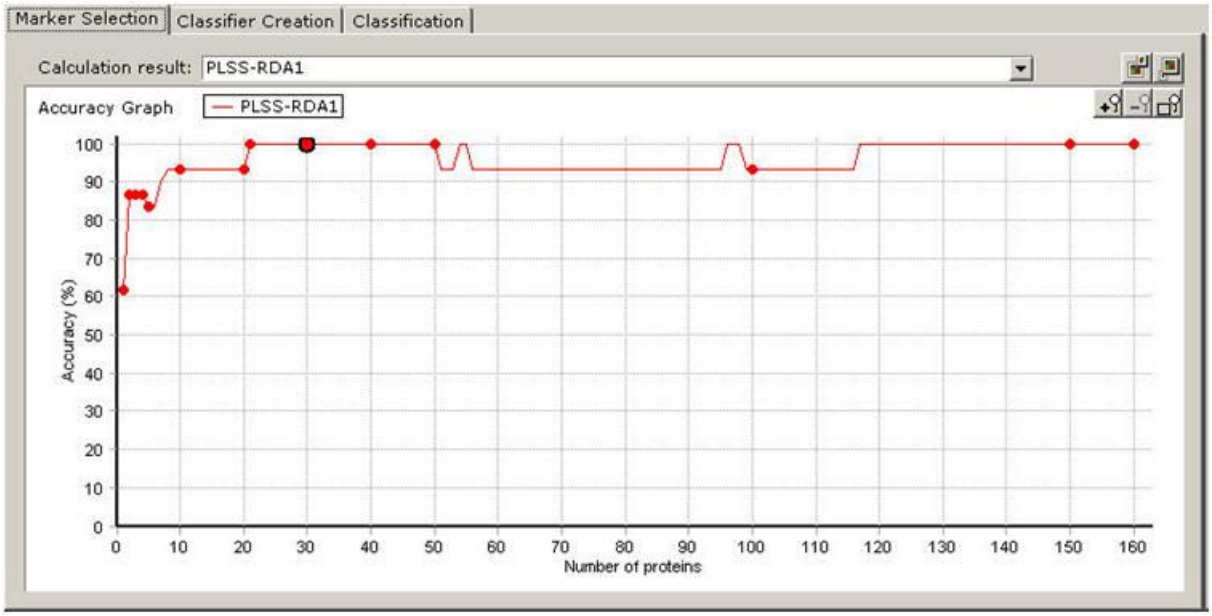


The left graph shows a differentially expressed protein, the right graph shows a non-differentially expressed protein. Because the data set contains a high number of non-differentially expressed proteins (noise), the results of the One-Way ANOVA will be used to filter out the differentially expressed ones (new set creation).

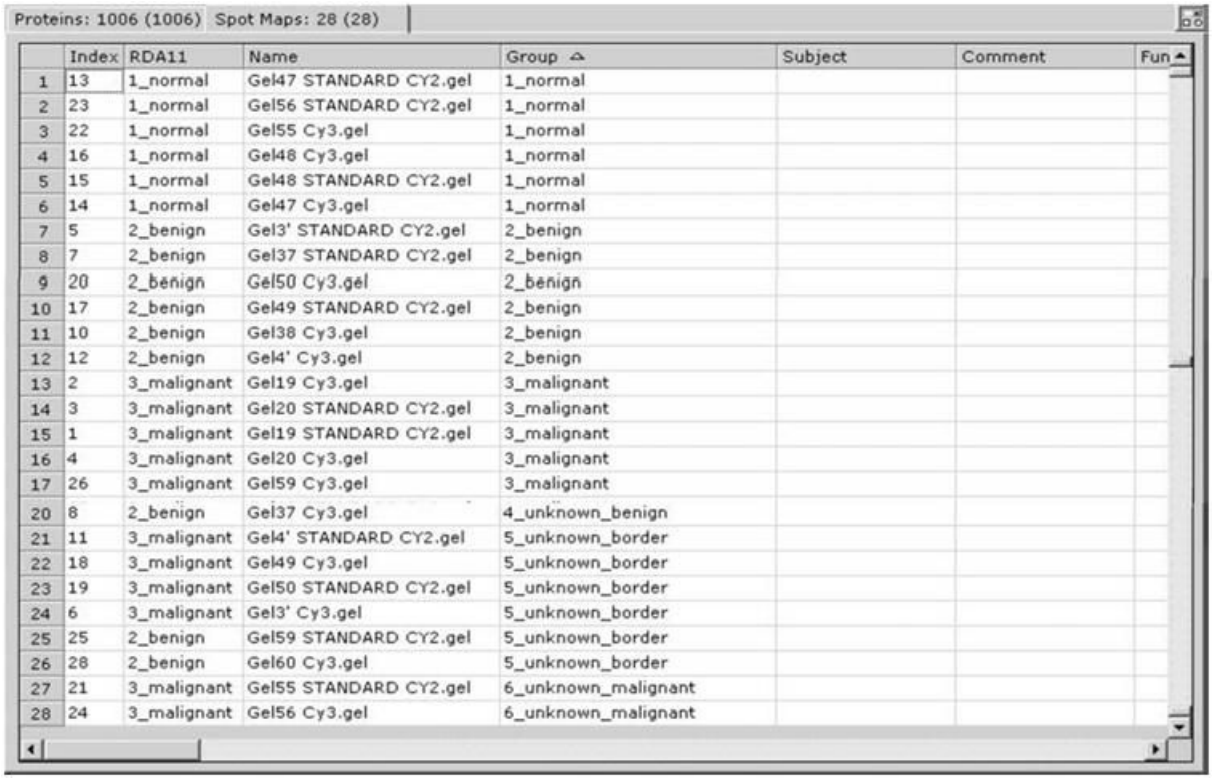
Marker selection

The Marker Selection analysis can be used to find a set of proteins that can be used to discriminate between experimental groups, e.g. benign tumors and malignant tumors.

If such a set is found, it is possible to create a classifier specialized for discriminating between e.g. the benign tumors and malignant tumors experimental groups (Classifier Creation) and to classify unknown samples.



Classification of known and unknown samples



Proteins: 1006 (1006) Spot Maps: 28 (28)

	Index	RDA11	Name	Group	Subject	Comment	Fun
1	13	1_normal	Gel47 STANDARD CY2.gel	1_normal			
2	23	1_normal	Gel56 STANDARD CY2.gel	1_normal			
3	22	1_normal	Gel55 Cy3.gel	1_normal			
4	16	1_normal	Gel48 Cy3.gel	1_normal			
5	15	1_normal	Gel48 STANDARD CY2.gel	1_normal			
6	14	1_normal	Gel47 Cy3.gel	1_normal			
7	5	2_benign	Gel3' STANDARD CY2.gel	2_benign			
8	7	2_benign	Gel37 STANDARD CY2.gel	2_benign			
9	20	2_benign	Gel50 Cy3.gel	2_benign			
10	17	2_benign	Gel49 STANDARD CY2.gel	2_benign			
11	10	2_benign	Gel38 Cy3.gel	2_benign			
12	12	2_benign	Gel4' Cy3.gel	2_benign			
13	2	3_malignant	Gel19 Cy3.gel	3_malignant			
14	3	3_malignant	Gel20 STANDARD CY2.gel	3_malignant			
15	1	3_malignant	Gel19 STANDARD CY2.gel	3_malignant			
16	4	3_malignant	Gel20 Cy3.gel	3_malignant			
17	26	3_malignant	Gel59 Cy3.gel	3_malignant			
20	8	2_benign	Gel37 Cy3.gel	4_unknown_benign			
21	11	3_malignant	Gel4' STANDARD CY2.gel	5_unknown_border			
22	18	3_malignant	Gel49 Cy3.gel	5_unknown_border			
23	19	3_malignant	Gel50 STANDARD CY2.gel	5_unknown_border			
24	6	3_malignant	Gel3' Cy3.gel	5_unknown_border			
25	25	2_benign	Gel59 STANDARD CY2.gel	5_unknown_border			
26	28	2_benign	Gel60 Cy3.gel	5_unknown_border			
27	21	3_malignant	Gel55 STANDARD CY2.gel	6_unknown_malignant			
28	24	3_malignant	Gel56 Cy3.gel	6_unknown_malignant			

Microarray training session

I. Designing and performing microarray experiment
II. Troubleshooting hybridization and washing
III. Data analysis and web resources
IV. Custom array design

A **DNA microarray** is a multiplex technology used in molecular biology. It consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, called features, each containing picomoles (10^{-12} moles) of a specific DNA sequence, known as *probes* (or *reporters*). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA sample (called *target*) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel. Therefore arrays have dramatically accelerated many types of investigation.

Microarray-based comparative genomic hybridization (array CGH) is a technique to scan the genome for gains and losses of chromosomal material. This method has significantly high resolution and clinical yield. The limitation of traditional karyotyping is that even high resolution karytoypes are unreliable for detecting many known microdeletion syndromes which range from 3-5 Mb in size.

Array CGH (aCGH) differentially labels DNA from a test sample and a reference sample. Fluorescently labeled reactions are combined and hybridized to DNA substrates from chromosomally normal individuals. Gains and losses of the genome in the test sample relative to the control are measured.

The strengths of this method are that it allows for investigation of the whole genome and has a very high resolution - allowing for the determination of origin of unknown genetic material such as unbalanced rearrangements. The main weakness of aCGH is that it does not detect balanced rearrangements. An uncertainty of aCGH is copy number variation (CNV). CNVs have been found in phenotypically normal people and can occur at a high frequency in the general population. They are usually inherited and can take up a significant portion of the genome. It is estimated that there are many more CNVs than we expect and they are more common than previously thought. It is not yet clear how CNVs contribute to genetic variability.

I. Designing and performing microarray experiment

Design and perform the microarray experiment by this protocol:

180k– Agilent labelling

Startdate experiment: …27.06.2011 ……

Performer experiment: …….

Date analysis: ……29.06.2011…….

Agilent 4x 180K oligo array; Manual protocol, Agilent labelling

Day 1 – Enzymatic digestion

	CME No.	Patient Name	[DNA] ng µl^{-1}	$^{260}/_{280}$ $^{260}/_{230}$	1000ng DNA = X µl	H_2O 20.2 – X µl	
1	MI 271	X	99	1.8 2.01	11.1	9.1	Cy 5
2	MI 276	X	193	1.87 2.05	5.2	15	
3	MI 65	X	275	1.78 2.1	3.6	16.6	
4	MI 79	X	72	1.88 2.2	13.9	6.3	
a	Male control		44	1.78 2.09	20.2		Cy 3
b	Male control		44	1.78 2.09	20.2		
c	Male control		44	1.78 2.09	20.2		
d	Male control		44	1.78 2.09	20.2		

1. Put **1000ng DNA** in 0.2ml PCR tube, add **H_2O** to a final volume of **20.2 µl**.
2. Prepare the Digestion Master Mix: no. of samples: ___8__

H_2O	**2.0 µl**		**= ____18_____**
10X Buffer L	**2.6 µl**		**= ____23.4____**
BSA (10 µg/µl)	**0.2 µl**	x9	**= ____1.8____**
AluI (10U/µl)	**0.5 µl**		**= ____4.5_____**
RsaI (10U/µl)	**0.5 µl**		**= ____4.5_____**

3. Add **5.8 µl** Digestion Master Mix to each tube to final volume of **26 µl**.
4. Incubate at **37°C** on PCR machine for **2 hours,** than on **65°C for 20 minutes** to inactivate the enzyme.
5. Cool to **4°C** on PCR machine.
6. Check 2 **µl** of the digestion on 0.8% agarose gel. The digested fragment shoud be in range of 200-500 bp.

Day 1 – Labelling

7. Add **5 µl** of **Random Primers Solution** to each sample.
8. Incubate at **98°C** on PCR machine for **3 minutes**
9. Cool to **4°C** on PCR machine for **10 minutes.**
10. Add the following to each tube: no. of samples: __8___

H_2O	**2.0 µl**		**= ___16_____**
5X reaction buffer	**10.0 µl**		**= ___80_____**
10X dNTP Nucleotide Mix	**5.0 µl**	**x8**	**= ___40_____**
Cy3™-dCTP or Cy5™-dCTP	**3.0 µl**		**= ___24_____**
Exo-Klenow Fragment	**1.0 µl**		**= ____8_____**

11. Add **21 µl** mix to each tube to final volume of **50 µl**.
12. Mix gently, quick spin down, place in light-proof box @ 37°C (oven / PCR machine).
13. Incubate at **37°C** for 4hrs or O/N.

Day 1 – Purification of labelled DNA

AMICON clean-up colums.

1. Add **450 µl TE** to each tube. Mix and spin down.
2. Transfer each mixture to Amicon spin column in fresh collection tube(provided).
3. Spin at **13000 rpm** for **10 minutes** in microcentrufuge.
4. Pour away eluant and replace spin column in the same collection tube.
5. Add **450 µl TE** to each spin column and spin at **13000 rpm** for **10 minutes**.
6. Invert the spin column and transfer to fresh, labelled collection tube (provided).
7. Spin at **1000g (3300 rpm)** for **2 minutes** to eluate the clean labelled DNA.
8. The eluated DNA should be **21 µl.** If is not, add TE to total volume of **21 µl.**
9. Record nanodrop measurement below:

	CME No.	Yield (µg)	[DNA] ng μl^{-1}	Cy 5 pmol (650nm)/SA	Cy 3 pmol (550nm)/SA	
1	MI 271	4.70	224	7.8/36.9		Cy 5
2	MI 276	6.18	294.5	109/48.6		
3	MI 65	5.98	285	9.7/34		
4	MI 79	5.33	254	8.5/33.5		
a	Male control	6.49	309		13.3/43	Cy 3
b	Male control	5.98	285		11.9/41.8	
c	Male control	5.10	243		10.1/41.6	
d	Male control	5.18	246		11.5/46.7	

Prepare the hybridization samples:

a) Equilibrate water baths or heat blocks to **95°C**.
b) Mix Cy-5 and Cy-3 labelled DNA in one tube.

MIX	A	B	C	D
Cy5	1	2	3	4
Cy3	a	b	c	d

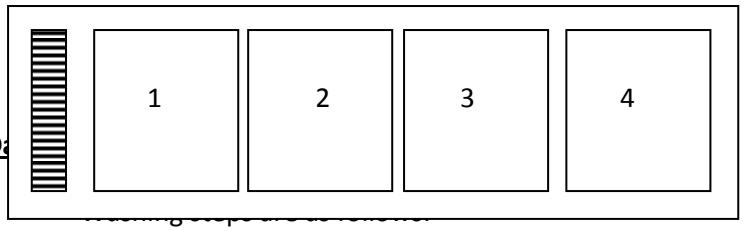
c) Add the following in the order indicated in new sample tubes:

Labeled gDNA mixture	**38 µl** (19µl Cy5 & 19µl Cy3)
Salmon Sperm DNA (0.9 mg/µL)	**5.6 µl**
Agilent Blocking Agent*	**11 µl**
Agilent 2X Hybridization Buffer	**55 µl**
Final Sample Volume	**~110 *µl***

*dilute lyophilised blocking agent in 1350µl H_2O (store @ -20°C)

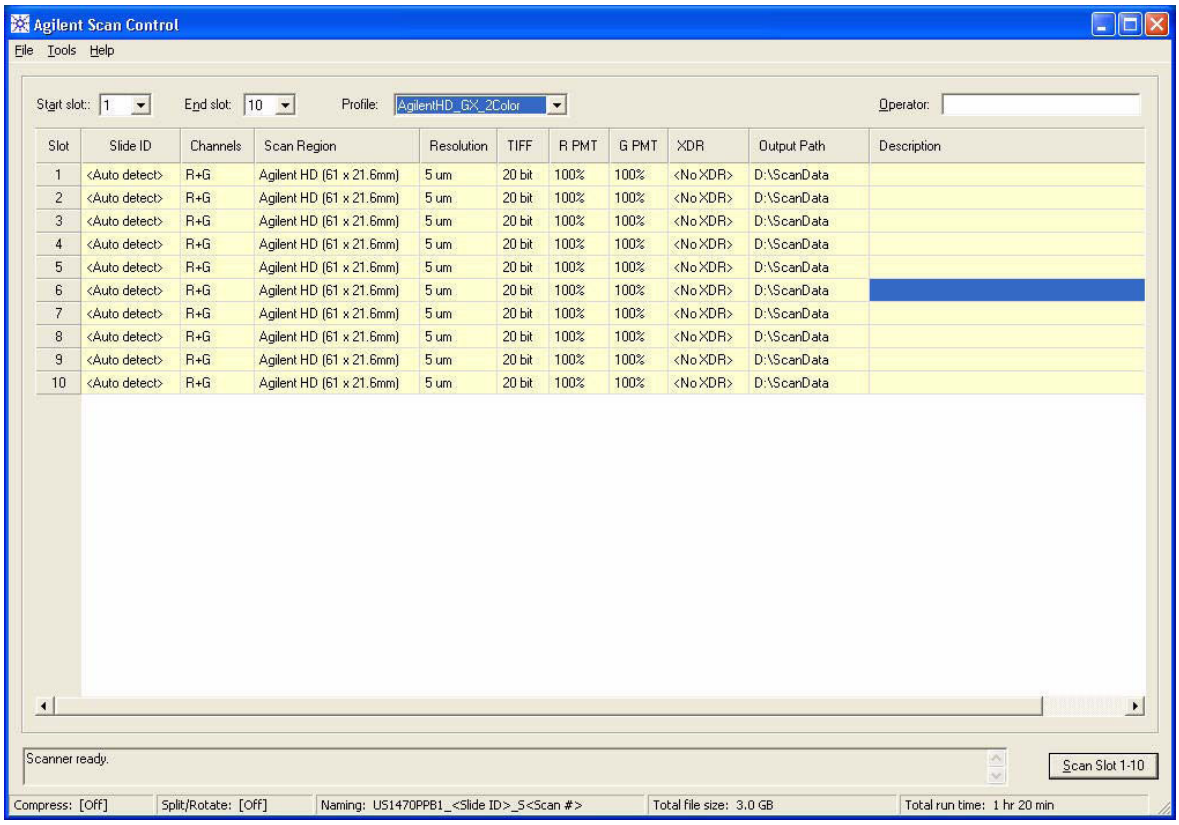
d) Mix by pipetting, quickspin
e) Transfer sample tubes to a water bath @ **95°C** for **10 minutes**.
f) Immediately transfer sample tubes to oven @ **37°C** for at least **30 minutes**.
g) Quickspin to collect the sample at the bottom of the tube.
h) Load **100 µl** of each hyb mix to appropriate gasket slide in Agilent chamber
i) Place Array on top of gasket slide with 'Agilent' facing down (slide no. facing up)
j) Close chamber tightly, place in Agilent oven @ 65°C, with rotation set to 20rpm
k) Hybridise for 40 hours

Agilent slide number		Cy5 sample	Cy3 sample
252206015955	[A] _1	MI 286	Male control
	[B] _2	MI 65	Male control
	[C] _3	MI 271	Male control
	[D] _4	MI 79	Male control

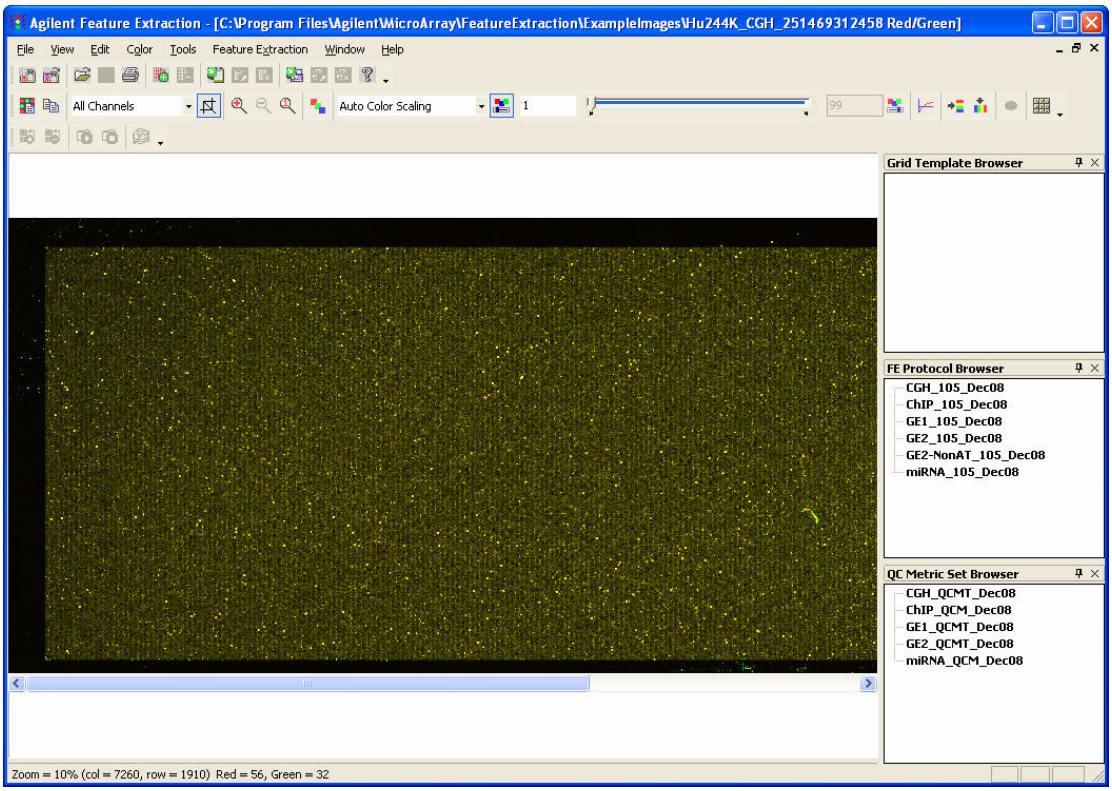


	Dish	Wash Buffer	Temperature	Time
Disassembly	#1	Oligo aCGH Wash Buffer 1	Room	
1st Wash	#2	Oligo aCGH Wash Buffer 1	Room	5min
2nd Wash	#3	Oligo aCGH Wash Buffer2	37˚C	1min

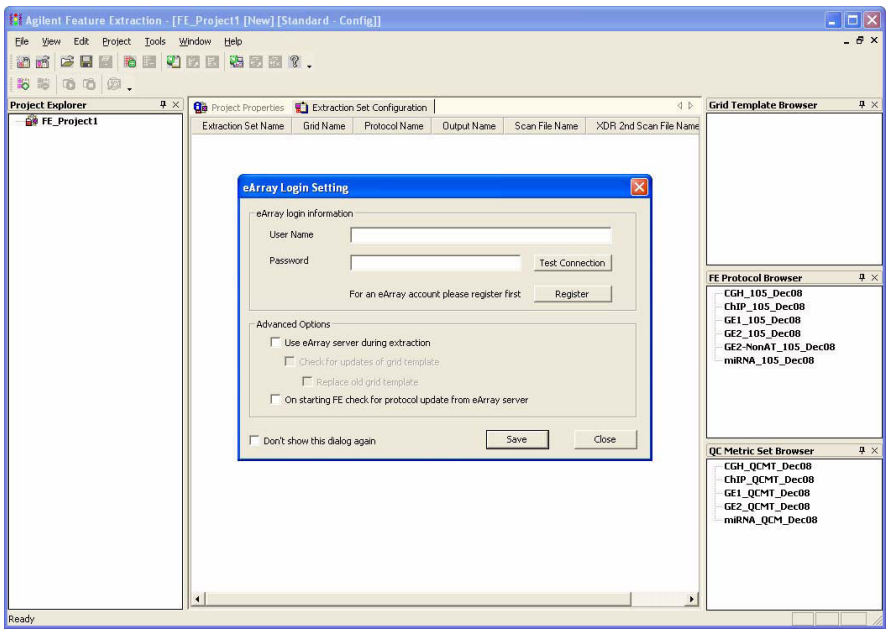
Washed slides are placed in slide holders, with ozone barriers. Scanning is performed in Agilent Scanner C. Properly placed holders into the slot of the scanner carousel are scanned at 3µm resolution, on red and green channel. This is the scan window when the scanner is recognized by the computer.



The scanner gives TIFF images of scanned microarray slides that are extracted by Feature extraction software v10.5.



The slide grid is recognized by the previously entered grid template. For every scanned TIFF image a new FE project is created.

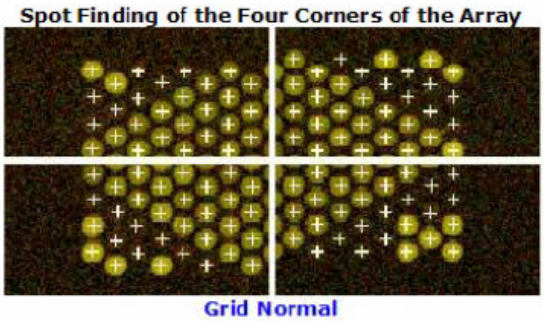


The FE software generates the QC (quality control) report file; JPEG and TEXT file which is uploaded into Genomic Workbench software.

The QC report contains values for the metrics shown bellow:

1.

Spot Finding of Four Corners
Look for:
Properly placed grid (grid normal) where the crosses align with the features

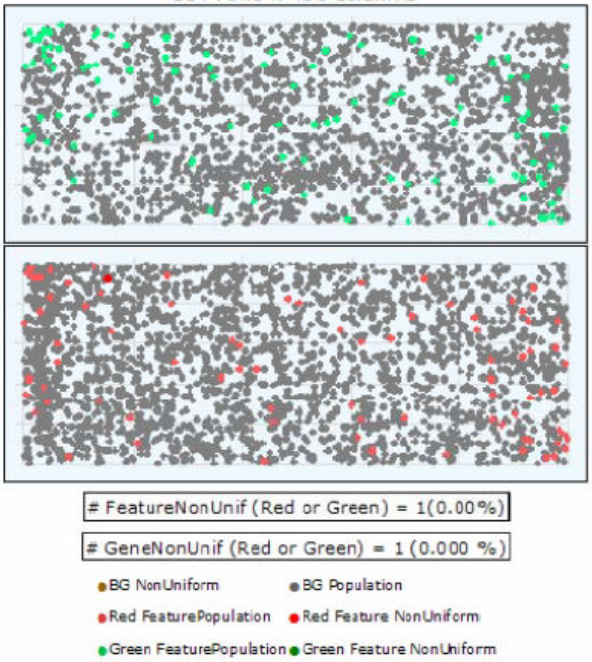


2.

Outlier Numbers with Spatial Distribution
Look for: A small number of Feature and Population non-uniform outliers that are randomly distributed on the array and not clustered

	Feature		Local Background	
	Red	Green	Red	Green
Non Uniform	1	0	0	0
Population	83	107	3170	2663

Spatial Distribution of All Outliers on the Array
534 rows x 456 columns



3.

Net Signal Statistics

Non-Control probes:

	Red	Green
# Saturated Features	0	0
99% of Sig. Distrib.	1083	697
50% of Sig. Distrib.	313	172
1% of Sig. Distrib.	101	41

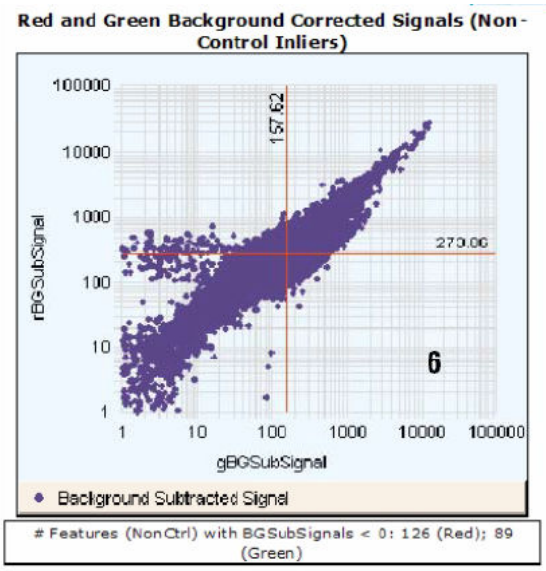
Negative Control Stats

	Red	Green
Average Net Signals	39.58	14.68
StdDev Net Signals	6.58	2.04
Average BG Sub Signal	0.76	0.33
StdDev BG Sub Signal (BG Noise)	5.78	2.01

Negative Control Stats – Average Net Signal:

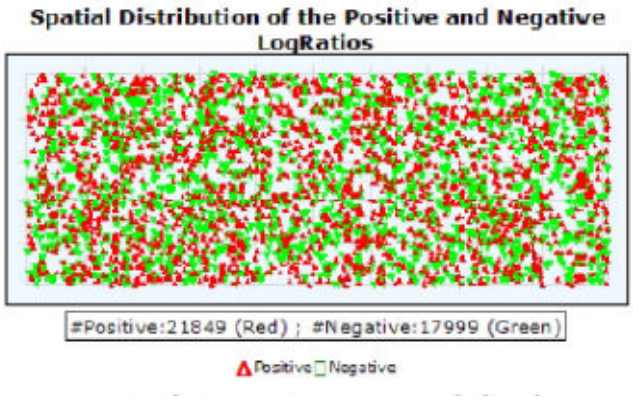
Negative control are structural control. High signal will highlight Background problems

4.



Plot of Background-Corrected Signals
Look for:
The majority of the data to be along the 45 degree angle with aberrations falling outside. High quality data has good correlation

5.



Spatial Distribution of Positive and Negative Log Ratios
Look for:
A random distribution across the array (note: these are not aberration calls, use DNA Analytics for calling aberrations). The number will depend on the sample.

QC report from array with good quality is shown bellow.

QC Report - Agilent Technologies : 2 Color CGH

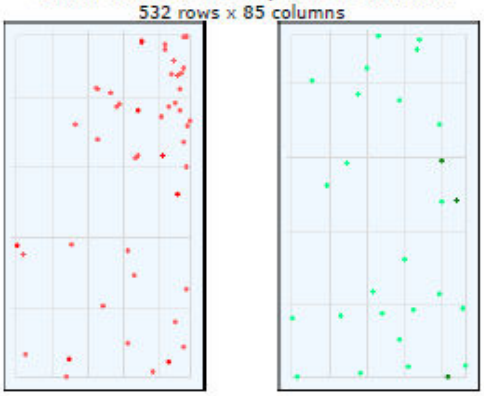
Date	June 29, 2011 13:30	Sample(red/green)	
User Name	Administrator	FE Version	10.5.1.1
Image	US92203694_251495023500_S01 [1_3]	BG Method	Detrend on (NegC)
Protocol	CGH_105_Dec08 (Read Only)	Multiplicative Detrend	True
Grid	014950_D_20070820	Dye Norm	Linear

Spot Finding of the Four Corners of the Array

Grid Normal

Outlier Numbers with Spatial Distribution

532 rows x 85 columns



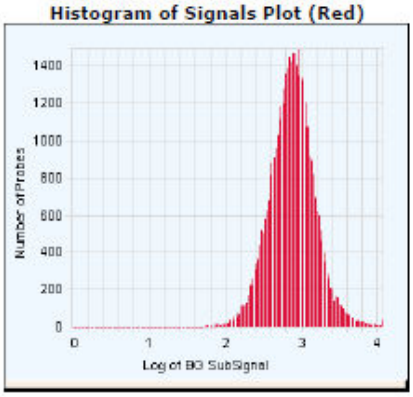
Red FeaturePopulation · Red Feature NonUniform · Green FeaturePopulation · Green Feature NonUniform

Feature	Red	Green	Any	% Outlier
Non Uniform	7	3	7	0.02
Population	43	24	65	0.14

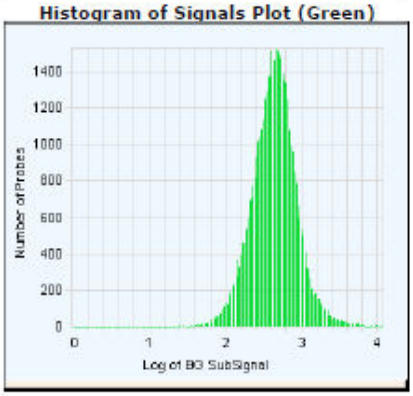
CGH_QCMT_Dec08 : (10 of 10) QCMetrics InRange

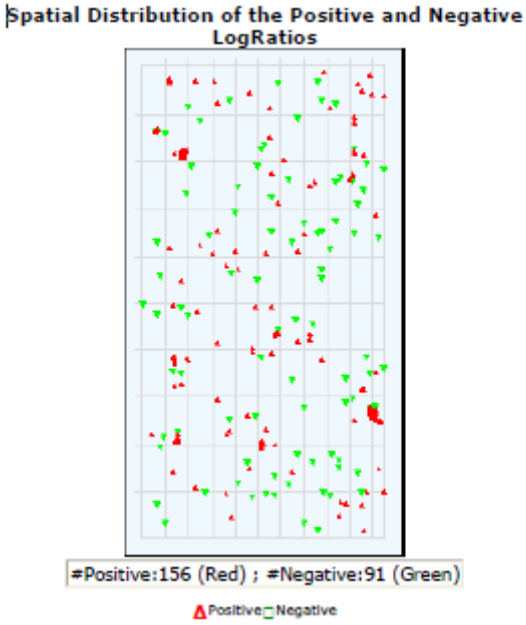
Metric Name	Value	UpLim	LowLim
AnyColorPrcntFeatNonUnifOL	0.02	1.00	NA
DerivativeLR_Spread	0.12	0.30	NA
gRepro	0.09	0.20	NA
g_BGNoise	7.50	15.00	NA
g_Signal2Noise	51.63	NA	30.00
g_SignalIntensity	387.40	NA	50.00
rRepro	0.09	0.20	NA
r_BGNoise	11.25	15.00	NA
r_Signal2Noise	58.31	NA	30.00
r_SignalIntensity	655.84	NA	50.00

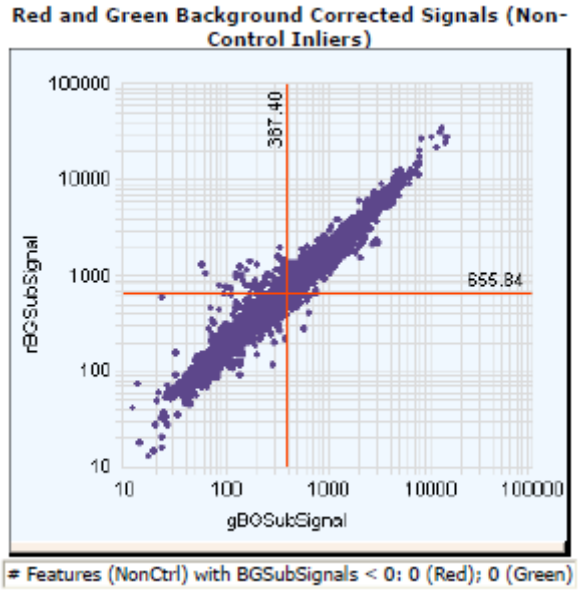
Histogram of Signals Plot (Red)



Histogram of Signals Plot (Green)




Spatial Distribution of the Positive and Negative LogRatios
#Positive:156 (Red) ; #Negative:91 (Green)
Positive
Negative


Red and Green Background Corrected Signals (Non-Control Inliers)
100000
10000
1000
100
10
rBGSubSignal
387.40
655.84
10
100
1000
10000
100000
gBGSubSignal
Features (NonCtrl) with BGSubSignals < 0: 0 (Red); 0 (Green)

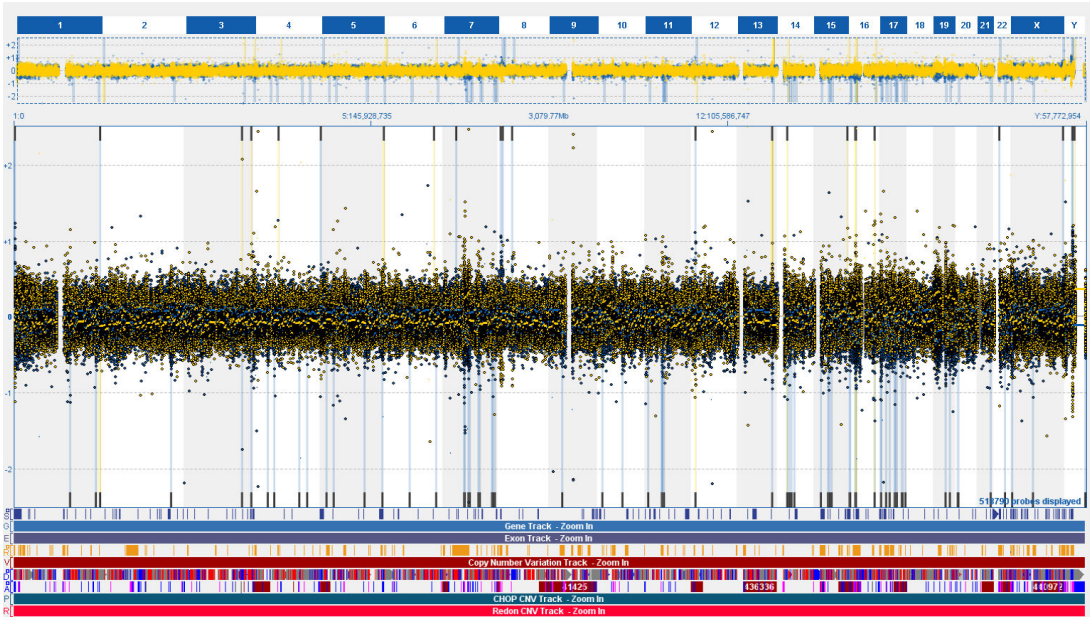
II. Troubleshooting hybridization and washing

1. High background noise values

High BGNoise can cause lower signal to noise values and higher DLRSD values. BGNoise is defined as the standard deviation of the signals on the negative controls. Wave effect calculation is recommended, and for each autosome:

- Calculate the moving average of the normalised log ratios using a 2Mb window size.
- For each band in the chromosome, calculate the standard deviation of the moving average for each probe in the band
- Calculate the mean of all of the standard deviations.
- Return the mean of this value across each autosome.
- A value of greater than 0.025 is considered poor

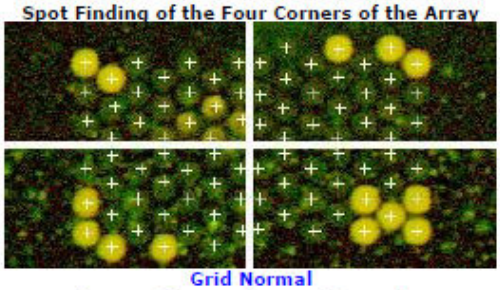
Wave effect – whole genome



QC Report - Agilent Technologies : 2 Color CGH

Date	Monday, December 20, 2010 - 15:06	Sample(red/green)	
User Name	Administrator	FE Version	10.5.1.1
Image	US92203694_252206015925_S01 [1_1]	BG Method	Detrend on (NegC)
Protocol	CGH_105_Dec08 (Read Only)	Multiplicative Detrend	True
Grid	022060_D_F_20081218	Dye Norm	Linear

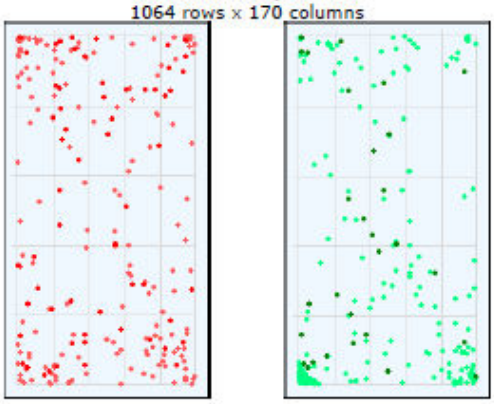
Spot Finding of the Four Corners of the Array



Grid Normal

Outlier Numbers with Spatial Distribution

1064 rows x 170 columns



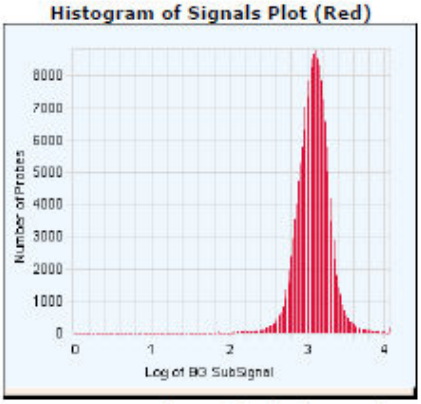
Red FeaturePopulation · Red Feature NonUniform · Green FeaturePopulation · Green Feature NonUniform

Feature	Red	Green	Any	% Outlier
Non Uniform	73	33	89	0.05
Population	242	239	417	0.23

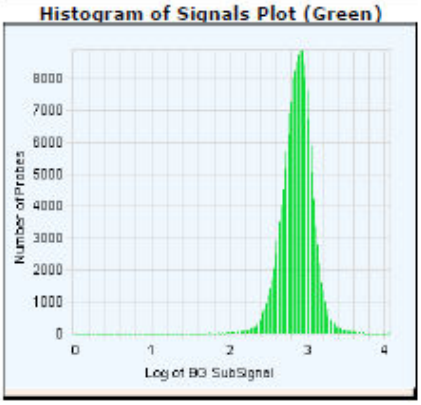
CGH_QCMT_Dec08 : (9 of 10) QCMetrics InRange

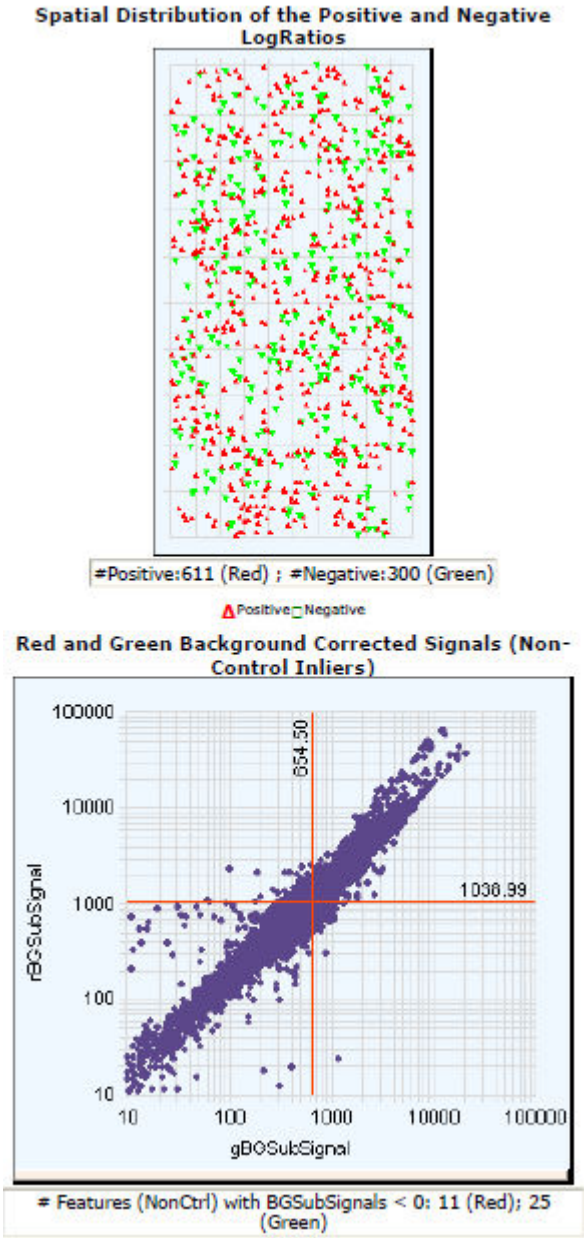
Metric Name	Value	UpLim	LowLim
AnyColorPrcntFeatNonUnifOL	0.05	1.00	NA
DerivativeLR_Spread	0.13	0.30	NA
gRepro	0.08	0.20	NA
g_BGNoise	13.37	15.00	NA
g_Signal2Noise	48.95	NA	30.00
g_SignalIntensity	654.50	NA	50.00
rRepro	0.08	0.20	NA
r_BGNoise	18.22	15.00	NA
r_Signal2Noise	57.04	NA	30.00
r_SignalIntensity	1038.99	NA	50.00

Histogram of Signals Plot (Red)



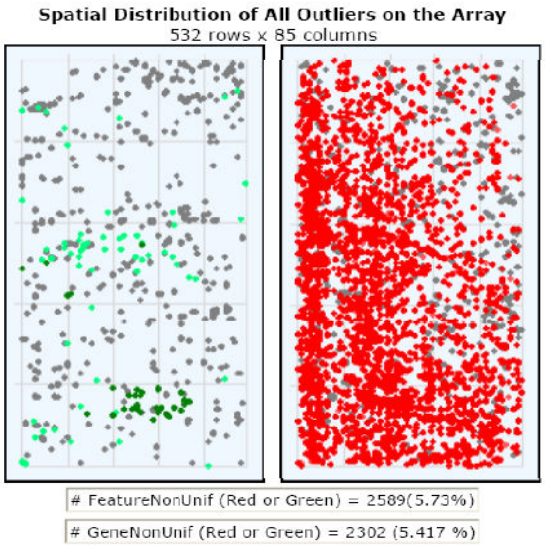
Histogram of Signals Plot (Green)



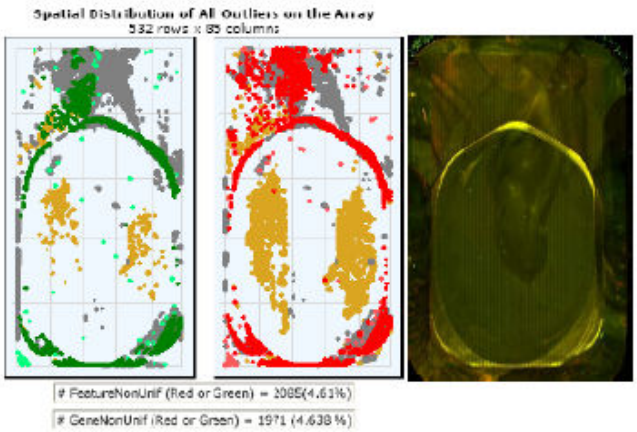


If the BGNoise is high examine the array image for visible non-uniformities. High BGNoise is often introduced during the washes. Try acetonitrile washing procedure.

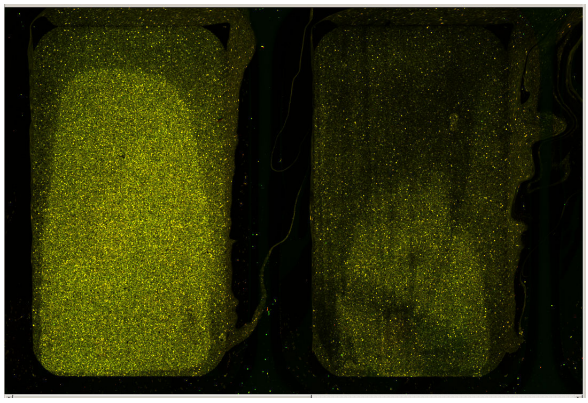
2. Ozone Problems



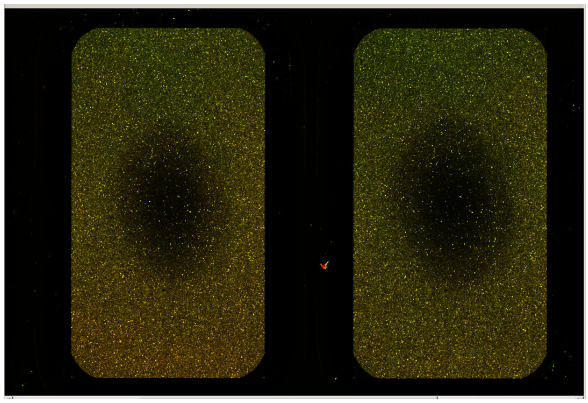
3. Hybridization Artifacts



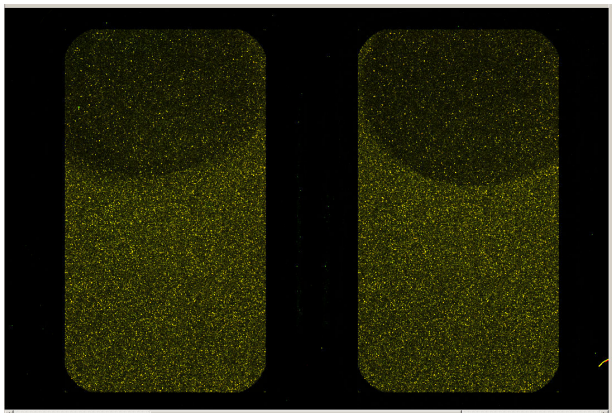
- Bad hybridisation - Leakage



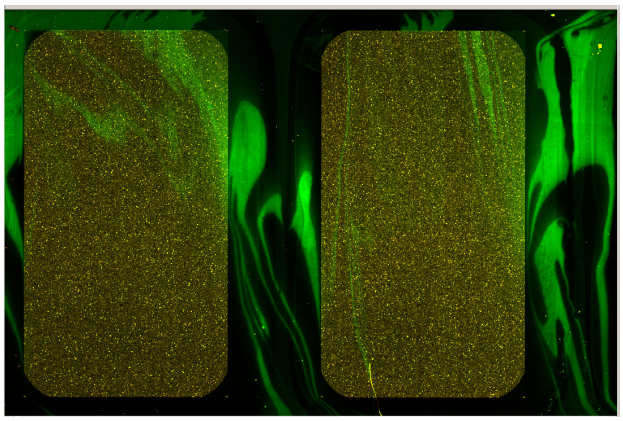
- Air bubbles too big



- Door open for too long - Rotation stopped / temperature dropped



4. Washing Artifacts

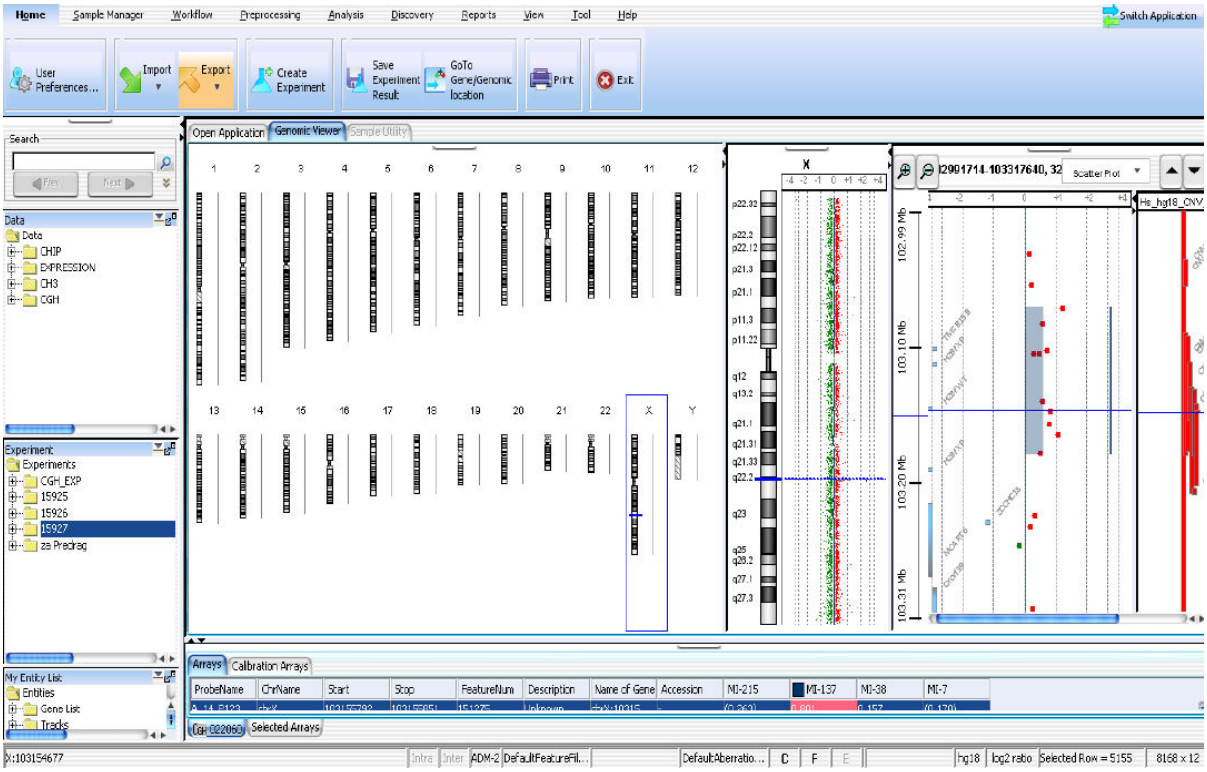


Troubleshooting

Symptoms	***Potential problem***	***Potential solution(s)***
Low Cy5 signals	Environmental conditions including ozone, humidity and temperature	Working in controlled environment Addition of antioxidants to hybridisation buffers and washing solutions Use of ozone stable fluorescent dyes (Voet and Vermeesch, 2007)
High SD	Low DNA quality	Control of DNA quality
Low SD	Insufficient suppression of repeat sequences	Control of Cot1 DNA quality
Increased background	Inadequate washing conditions	Adjustment of washing conditions
Fluorescence signal heterogeneity along the array	- Poor laser adjustment - Ozone/light degradation - Array quality	Normalization by sub-array (block-normalization) Adjust laser, proper maintenance of appliance
Artefactual ratio changes – labelling bias	Molecular structure of Cy3 and Cy5	Indirect labelling approaches using the incorporation of aminoalyl modified dNTPs into both DNAs, followed by independent direct labelling with reactive fluorochromes. Introduction of replicate dye-reversal hybridizations or direct chemical labelling by using ULS- coupled dyes
Uneven hybridisation	Manual hybridisation with drying/leaking	Automated hybridization process
Low intensity, high SD	Low incorporation, poor labelling efficiency, poor recovery	- Strong QC for DNA quality and labelling - Measure specific activity (typically 1 fluorophore every 30-80 bp)
Low dynamic range	- Bad quality arrays - Washing not stringent enough - spot saturation	- BAC spots are contaminated - Check conditions

III. Data analysis and web resources

Genomic Workbench software (Agillent) is used for data analysis. The Agilent Genomic Workbench provides a powerful visualization tool within a seamless environment for the design, analysis and management of key microarray applications. Use Agilent Genomic Workbench 6.0 User Guide.



Database of Genomic Variants

The objective of the Database of Genomic Variants is to provide a comprehensive summary of structural variation in the human genome. The content of the database is only representing structural variation identified in healthy control samples.The Database of Genomic Variants provides a useful catalog of control data for studies aiming to correlate genomic variation with phenotypic data.

web page: http://projects.tcag.ca/variation/

NCBI Map Viewer

The NCBI Map Viewer provides graphical displays of features on the human reference genome sequence assembly maintained by the GRC and the alternate HuRef genome assembly, as well as cytogenetic, genetic, physical, and radiation hybrid maps. Map features that can be seen along the sequence include genes, transcripts, NCBI contigs (the 'Contig'

map), the BAC tiling path (the 'Component' map), STSs, FISH mapped clones, ESTs and transcripts from several different organisms, Gnomon predicted gene models, and more.

web page: http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606

DECIPHER

DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources) is an interactive web-based database which incorporates a suite of tools designed to aid the interpretation of submicroscopic chromosomal imbalance. Many patients suffering from developmental disorders harbour submicroscopic deletions or duplications that, by affecting the copy number of dosage-sensitive genes or disrupting normal gene expression, lead to disease. However, many aberrations are novel or extremely rare, making clinical interpretation problematic and genotype-phenotype correlations uncertain. Identification of patients sharing a genomic rearrangement and having phenotypic features in common leads to greater certainty in the pathogenic nature of the rearrangement and enables new syndromes to be defined.

web page: http://decipher.sanger.ac.uk/

CHOP Database

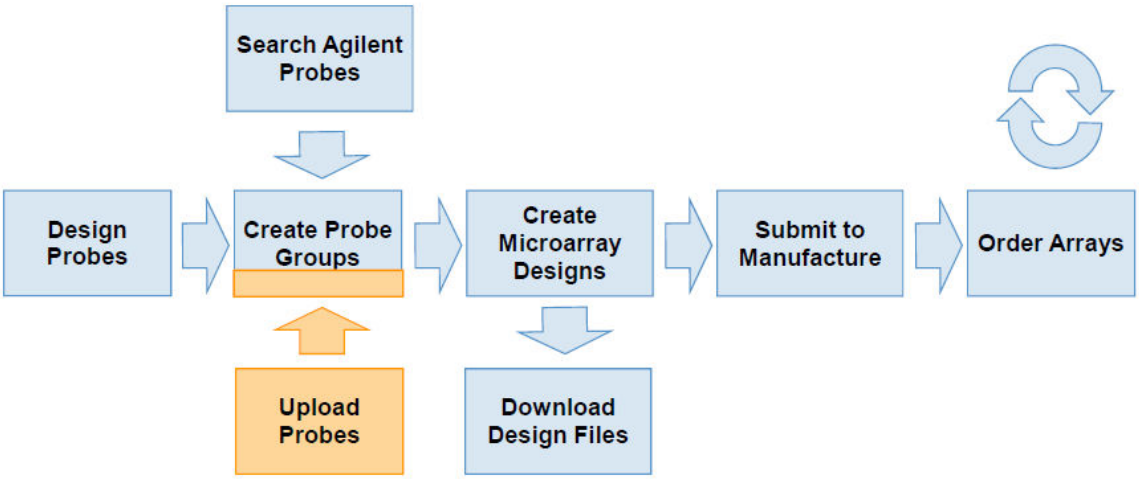
Researchers from the Children's Hospital of Philadelphia and the University of Pennsylvania have released a database of copy number variations present in thousands of healthy individuals. The resource, called the Copy Number Variation Project, joins a growing list of CNV databases, including the Data of Genomic Variants hosted by the Hospital for Sick Children in Toronto and the Wellcome Trust Sanger Institute's DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources).

web page: http://www.genomeweb.com/arrays/chop-develops-database-cnvs-healthy-individuals-use-controls-future-studies

IV. Custom array design

For specific research purposes there is a tool provided by Agilent for designing custom microarrays. eArray is a web-based application http://earray.chem.agilent.com that allows researchers to choose the probes they need either from their database (oligo libraries) or from the earray catalog. eArray is free to all registered users, and allows to create assays that are specific to your research needs.

eArray provides an easy way of managing the array creation process.



The content of an Agilent microarray is defined as the union of the Probe Groups it contains. During the creation of a custom microarray content is added by associating it with one or more Probe Groups. These Probe Groups can either be preexisting or created on the fly using eArray's Design Wizards. A Probe Group consists of one or more probes assigned to a group. Probes can exist in more than one Probe Group. It may be useful to split regions (*e.g.*, regions designed to address different disorders or regions with different probe densities) into separate Probe Groups, considering possible mix-and-match adjustments in future array designs. If you want to have a set of probes that will be replicated or used as normalization probes then these need to be in a separate Probe Groups. Probe Groups can be created in multiple ways:

- (Strongly recommended) By searching the Agilent HD-CGH database, a high-density database of predesigned probes
- By uploading your own probes
- By searching your own probes or Probe Groups
- By 'Genomic Tiling', *i.e.,* designing probes at a precise spacing

Every probe in the Agilent HD-CGH Probe Database has a probe performance score with a value between 0 and 1. The higher the score, the greater the likelihood that a probe will produce a good log ratio response when it is included on an Agilent CGH microarray. The probe score generated by eArray is based on a statistical model that was constructed by regressing *in silico* parameters for the probe sequence against an observed log ratio response in a model system. The parameters used include Tm, GC content, a hairpin ΔG, sequence complexity, and metrics to measure homology with the rest of the reference genome. This probe score is a prediction of the response of the particular probe in log ratio space. The response is defi ned as observed over expected log ratio. The average probe

score for all probes in the HD database is 0.759, the average probe score for the Agilent catalog array AMADID 021529 (SurePrint G3 Human CGH Microarray 1x1 M) is 0.917. You can obtain scores from eArray for custom probes (*e.g.*, probes designed using the 'Genomic Tiling' function). To do this, submit a probe scoring job (Under "Probes", "Score Custom Probes"). During the custom probe scoring, all metrics described above are computed and plugged into the model to generate the probe score. Once the probe scoring job completes the CGH probe scores will be available in a download from the probe scoring job or by downloading the tdt fi le from the Probe Group. Low scoring probes can then be fi ltered outside of eArray by downloading the Probe Group with the scores and making a list of probe IDs that pass your score threshold. That list can be used to search probes in your eArray folder and create a new Probe Group. In DNA Analytics (part of Genomic Workbench 6.0 or higher), probes can be fi ltered by probe score if the probes used to create the design have been scored.

A control grid is a set of control probes that are added to every array design. For every Agilent microarray design format and species there is a default species specific control grid suggested in eArray. These grids contain positive controls probes designs against endogenous sequence. Positive controls can be used for image orientation and to assess whether the sample was labeled. The "genomic" control grids are generic grids that contain **only** negative controls. Since they lack endogenous positive control probes they can be used for any species. The negative control probes are used to measure on element background. Here are some of the probe types you should expect to fi nd on a species-specifi c genomic control grid.

Normalization or 'backbone' probes should be included on a CGH microarray when it is expected that most probes are going to be in aberrant regions, such as a very focused design where half of the probes are on the X-chromosome. When designing a whole-genome array, though, it is not necessary to include specifi c normalization probes since there should be a sufficient number of non-aberrant regions for proper normalization. If you choose to include normalization probes, the minimum number of normalization probes is one percent of the microarray (Feature Extraction requirement) and the recommended number is at least several hundred probes. Each microarray design format has a default Agilent normalization Probe Group associated with it.

Replicate probes are used to calculate the QC metric Reproducibility. This metric is set to the median percent CV of background-subtracted signal for these replicate probes after outlier rejection. High scores for the Reproducibility metric usually indicate that the hybridization volume was too low or that the oven stopped rotating during the hybridization. The minimum number of replicated probes should be >= 300 probes replicated fi ve times. This is because Feature Extraction requires a minimum three times replication level after rejection of feature non-uniformity outliers (FNUOL). The default Agilent replicate Probe Group for the 1x244 K and

1x1 M CGH microarrays contains 1000 probes and should be replicated fi ve times. The 1000 probes are a random selection from catalog arrays. (www.opengenomics.com/cgh www.opengenomics.com/earray).

Real-time PCR training session

IV. CNV validation by real-time PCR

V. Real-time PCR data analysis

Real-time PCR is a method for quantification of specific DNA sequences in complex mixtures of sequences. The real-time PCR technology provides a way to monitor the progress of the amplification as it occurs (in real time) circumventing the need for post-PCR analysis. Reactions are analysed by the point in time during cycling when the quantity of the specific PCR product reaches an arbitrary threshold. The earlier this happens, the higher the copy-number of the specific target prior to cycling.

The technology is based on measurements of fluorescence intensity whereby specific PCR amplification results in increased fluorescence of a given fluorescent dye. Thus, the amount of specific PCR product during the reaction is proportional to the fluorescence intensity for the specific fluorescent dye. A range of chemistries have been developed, such as TaqMan probe-based 5' nuclease chemistry, molecular beacon chemistry and various intercalating dye chemistries.

Intercalating dyes. A real-time PCR reaction exploiting intercalating dye chemistry contains primers, DNA template, thermostable polymerase, buffer, deoxynucleotides and a dye that fluoresces efficiently only when complexed with double-stranded DNA. Examples of dyes with suitable properties supplied by different companies are SYBR Green I, SYTO9, EvaGreen. Typically, all components except for primers and DNA, are provided as a pre-mixed solution that needs to be diluted down to the optimal working concentration of its components.

Standard curve method. Accurate quantification of a specific sequence in a sample of interest can be achieved by amplifying this sequence in the sample of interest in parallel to amplification of the same sequence in a number of standards. The standards should contain known numbers of copies of the sequence of interest. The standards serve to build a standard curve and the unknown sample is quantified against the standard curve.

Rationale of the experiment. aCGH experiments performed in the laboratory have identified a number of copy-number variations (CNVs) in male infertility patients. These results are to be confirmed by an independent method, i.e. real-time PCR. Primer pairs have been designed within two CNVs of interest on chromosomes 7 and X. DNA isolated from the patients is to be quantified against a standard curve prepared from DNA from a control individual that has been shown not to carry CNVs in the regions of interest. In addition, a separate control gene, *B2M*, is also amplified by specific primers. Neither the patients, nor the control individual carry any CNVs in this region. This normaliser amplicon is used to control for small differences in DNA concentration between individuals.

Experimental procedures – plate design and sample preparation

→ Design your experiment. Select your primers, unknown samples, standards and other reagents:

PCR primers: human primer pairs: 7_q31.1_b_FOR/7_q31.1_b_REV;
X_q28_a_FOR/X_q28_a_REV;
X_q28_b_FOR/X_q28_b_REV;
B2M_FOR/B2M_REV

Samples: DNA from patients with male infertility: MI-7, MI-38, MI-88, MI-91, MI-100, MI-137, MI-215: 5 ng/µl tubes

Standards: DNA from control individual PN, 100 ng/µl

Real-time PCR master mix: 5 × HOT FIREPol® EvaGreen® qPCR Mix Plus (Solis BioDyne, Cat. No. 08-24-00020)

Real-time PCR machine: 7500 Fast Real-Time PCR System
PCR cycling conditions have been optimised in previous experiments.

Dilutions

Standards:

S 100: DNA PN 100 ng/µl	concentration 100 ng/µl
S 25: : 75 µl water + 25 µl PN 100	concentration 25 ng/µl
S 6.25: 75 µl water + 25 µl PN 25	concentration 6.25 ng/µl
S 1.56: 75 µl water + 25 µl PN 6.25	concentration 1.56 ng/µl
S 0.39: 75 µl water + 25 µl PN 1.56	concentration 0.39 ng/µl

1 µM each primer working stocks: 196 µl H_2O + 2.0 µl 100 µM forward primer + 2.0 µl 100 µM reverse primer (1 µl/10 µl reaction, final 100 nM)

Reaction mixtures:

H_2O	6.0 µl	
5 × EvaGreen mix	2.0 µl	(to 1 ×)
Template	1.0 µl	(variable)
1 µM primer stock	1.0 µl	(100 nM each)
final volume	10 µl	

Master mix for 28 reactions (one per each primer pair)

H_2O	168 µl
5 × EvaGreen mix	56 µl
1 µM primer stock	28 µl

Distribute 9 µl of master mix per reaction – see table below.

	1	2	3	4	5	6	7	8	9	10	11	12
A	7_q31.1_b_FOR/7_q31.1_b_REV											
B												
C	X_q28_a_FOR/X_q28_a_REV											
D												
E	X_q28_b_FOR/X_q28_b_REV											
F												
G	B2M_FOR/B2M_REV											
H												

To each well add 1.0 µl template or 1.0 µl water to negative controls – see table below.

	1	2	3	4	5	6	7	8	9	10	11	12
A	S 100	S 100	S 25	S 25	S 6.25	S 6.25	S 1.56	S 1.56	S 0.39	S 0.39	MI-7	MI-7
B	MI-38	MI-38	MI-88	MI-88	MI-91	MI-91	MI-100	MI-137	MI-137	MI-215 *	MI-215 *	NTC
C	S 100	S 100	S 25	S 25	S 6.25	S 6.25	S 1.56	S 1.56	S 0.39	S 0.39	MI-7 *	MI-7 *
D	MI-38	MI-38	MI-88	MI-88	MI-91	MI-91	MI-100	MI-137	MI-137	MI-215	MI-215	NTC
E	S 100	S 100	S 25	S 25	S 6.25	S 6.25	S 1.56	S 1.56	S 0.39	S 0.39	MI-7 *	MI-7 *
F	MI-38	MI-38	MI-88	MI-88	MI-91	MI-91	MI-100	MI-137	MI-137	MI-215	MI-215	NTC
G	S 100	S 100	S 25	S 25	S 6.25	S 6.25	S 1.56	S 1.56	S 0.39	S 0.39	MI-7	MI-7
H	MI-38	MI-38	MI-88	MI-88	MI-91	MI-91	MI-100	MI-137	MI-137	MI-215	MI-215	NTC

→ Run the PCR on the real-time PCR instrument. In the 7500 real-time PCR system software, create an experiment or plate document using the following parameters:

Experiment Properties

Type of experiment: Quantitation – Relative Standard Curve

Type of reagents: SYBR Green® Reagents

Ramp speed: Standard

Plate Setup; Define Targets and Samples

Targets: 7_q31.1_b, X_q28_a, X_q28_b, B2M; assign to individual wells; define NTCs

Samples: MI-7, MI-38, MI-88, MI-91, MI-100, MI-137, MI-215; assign to individual wells

Run Method

Thermal cycling conditions: Enzyme activation and initial denaturation at 95°C for 15 min followed by 40 cycles: 95°C, 15 s; 60°C, 60 s
Standard melting conditions

* Carries CNV in this region by aCGH

Note: For more details refer to the Applied Biosystems 7500/7500 Fast Real-Time PCR System Getting Started Guide for Comparative CT/Relative Standard Curve Experiments (Part Number 4387783).

Analysis of the data

→ Review amplification curves on the ΔRn versus cycle number graph.

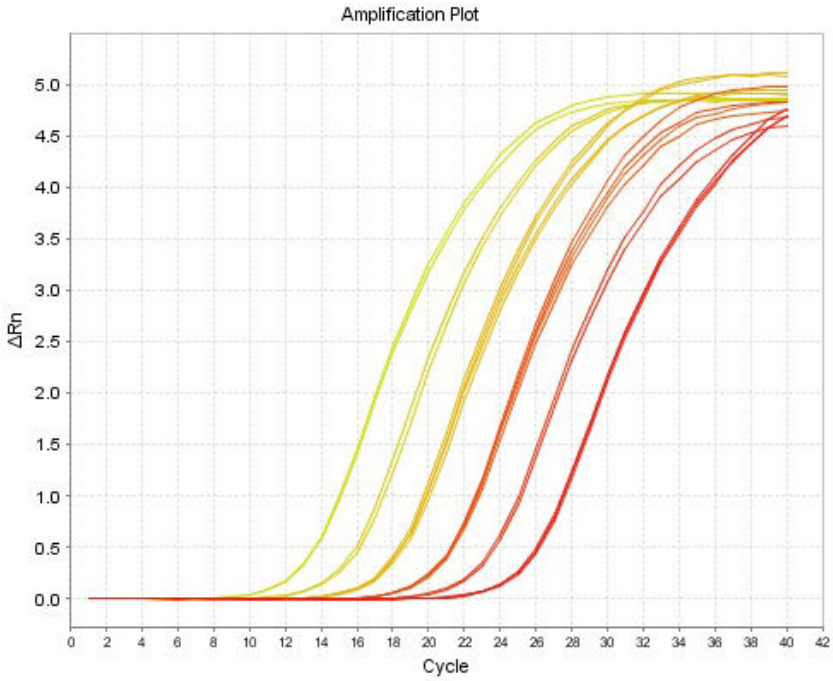


Figure 1. Graph representing fluorescence intensity (normalised, baseline values subtracted) versus cycle number for several reaction wells. The real-time PCR amplification curves are shown in linear scale.

The y axis can be linear (Figure 1) or logarithmic (Figure 2). The amplification in the earlier cycles becomes obvious when the graph is set to log scale. It is during these earlier cycles that the amplification is most efficient. It is in fact exponential since the amount of PCR product in the reaction doubles with each cycle. This exponential phase forms the linear part of the curve in log scale. It is followed by a plateau phase indicating reduced amplification efficiency due to the depletion of one or more reaction components. A threshold is automatically set within the linear range. The cycle number at which the curve for a specific reaction crosses the threshold is called the Ct value of the reaction.

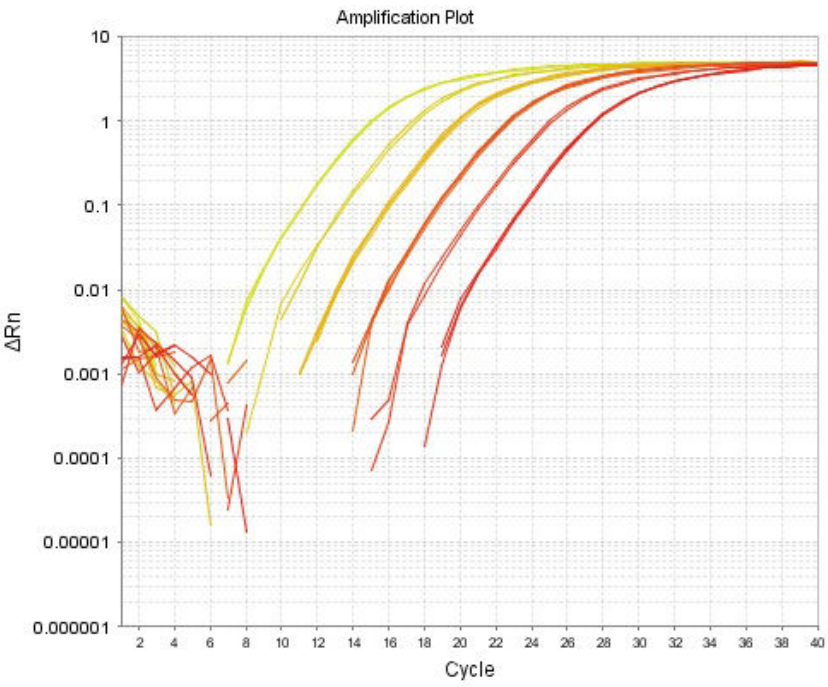


Figure 2. Graph representing fluorescence intensity (normalised, baseline values subtracted) versus cycle number for several reaction wells. The same real-time PCR amplification curves shown in Figure 1 are represented in log scale.

→ Review melting curves (Figure 3). Melting curves are produced post-amplification to reveal the complexity of the amplified sample.

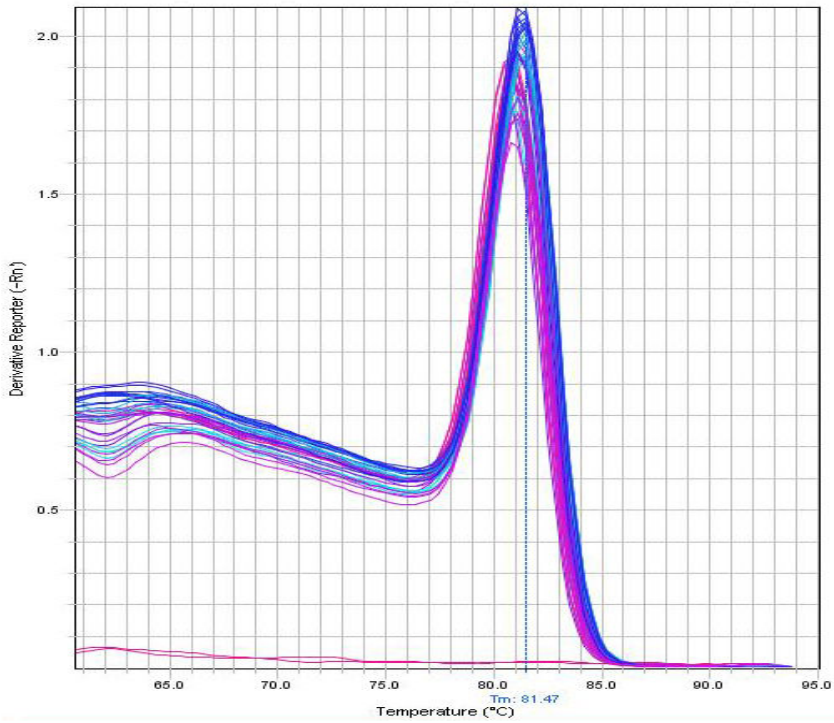


Figure 3. Melting curves reveal the specificity of amplification.

Specific amplification results in one melt peak of characteristic temperature. The presence two or more peaks is indicative of primer dimers or other undesirable amplification that has taken place during cycling. The fluorescence signal accounted for by non-specific amplification forms part of the total signal and therefore contributes to the Ct value of the reaction affecting the accuracy of the assay. When multiple peaks are observed it is necessary to optimise PCR conditions and/or redesign PCR primers.

→ Review the standard curves (Figure 4). The standard curves are produced by plotting the Ct values of the reactions containing dilutions of the standard DNA sample against DNA quantity. A standard curve has to be generated for each of the assays/primer pairs used in the experiment.

A good standard curve should have a correlation coefficient (R^2 value) of at least 0.990. Values of 0.998 and over can be achieved with well designed assays. In order to produce good quality standard curves, depending on the design of the plate, it is advisable to omit outlier wells from the analysis.

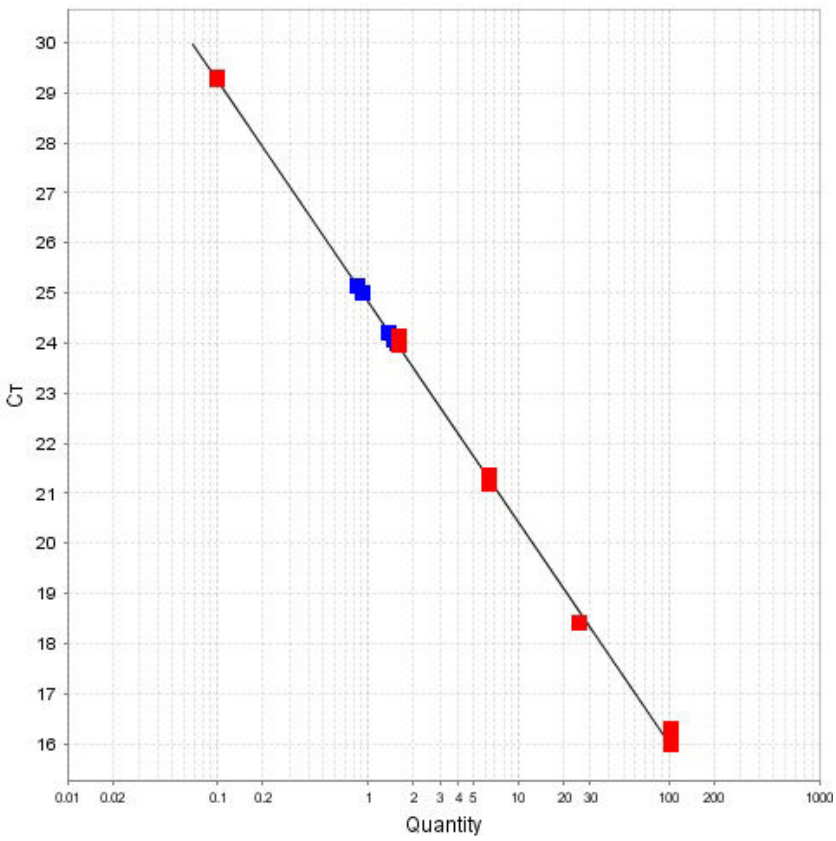


Figure 4. Standard curve for one real-time PCR assay/primer pair. Red squares depict individual reactions set up with known quantities of standard DNA, while blue squares represent unknown samples to be quantified.

→ Obtain quantities for unknown samples. Unknown samples are directly quantified using the standard curve. Thus, a measure of the quantity of the amplified sequence in the original sample is obtained. The quantity is expressed in the same units as the quantity of the standards, e. g. number of molecules per sample or equivalent. In a Relative Standard Curve experiment, the final results are independent of the choice of units used for quantification.

→ Normalise results using the normaliser assay/primer pair and present them as number of genomic copies carried by each patient. The normaliser assay is used to correct for differences in genomic DNA concentration between samples. Having in mind that the normaliser sequence is not deleted in any of the samples and that the control individual does not carry deletions or duplications in the region of interest, calculate the number of genomic copies of each of the amplified sequences for each patient.

Having in mind that the normaliser sequence is not deleted in any of the samples and that the control individual does not carry deletions or duplications in the region of interest, calculate the number of genomic copies of each of the amplified sequences for each patient.

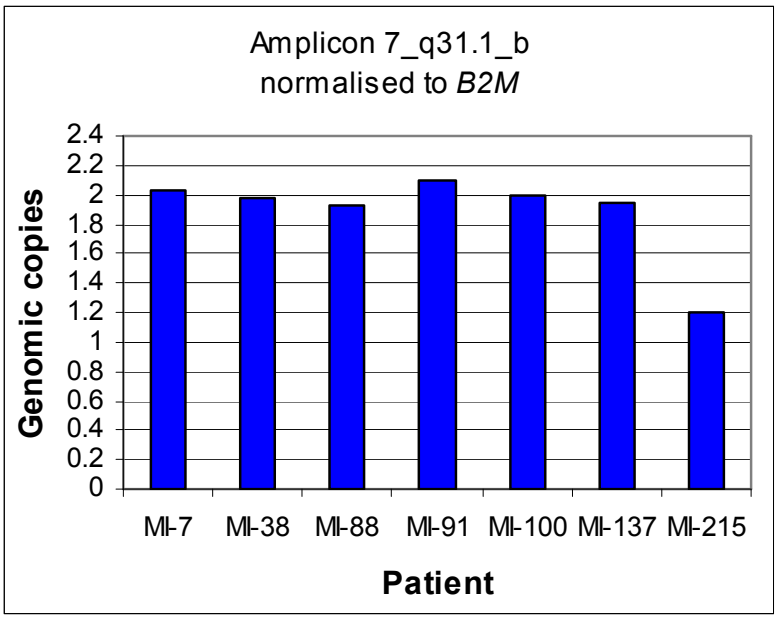


Figure 5. Expected results from the validation of CNVs by real-time PCR. Results for one assay/primer pair are shown.