

Análise de Softwares para Aquisição de Imagens em Experimentos de cDNA Microarrays

Aluno: Gustavo Henrique Esteves

Orientador: Luiz F. L. Reis - ILPC

Co-orientadores: Junior Barrera - IME - USP

Alex F. Carvalho - ILPC



Sumário

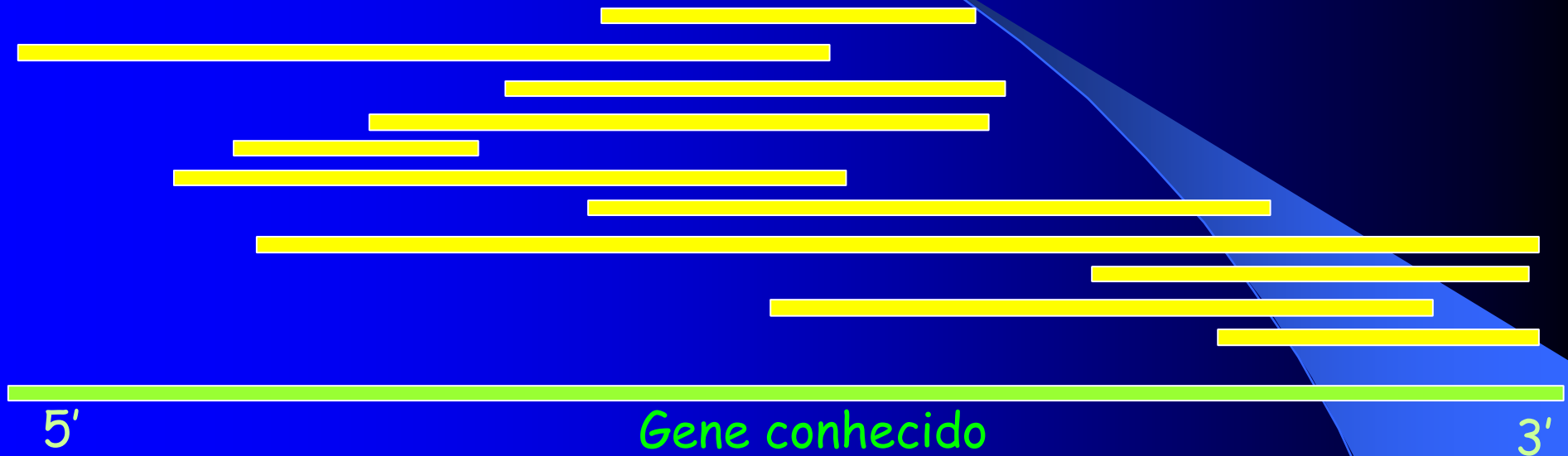
1. *A tecnologia de microarray*
2. *Medida dos dados*
3. *Objetivos do trabalho*
4. *Metodologia*
5. *Resultados preliminares*



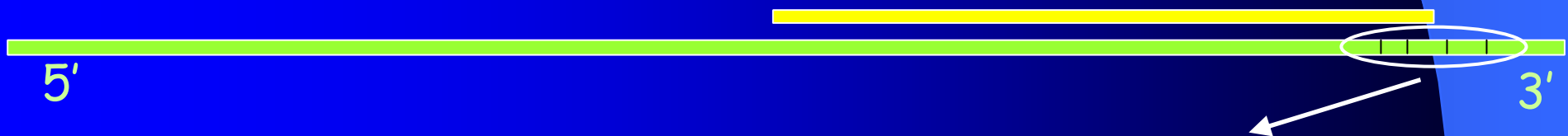
*A Tecnologia de
Microarrays*

Seleção dos clones de interesse

1. Criação de *clusters* do mesmo gene

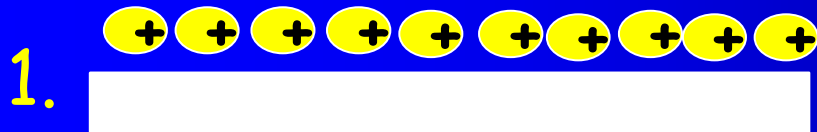


2. Escolha do clone que melhor represente o gene

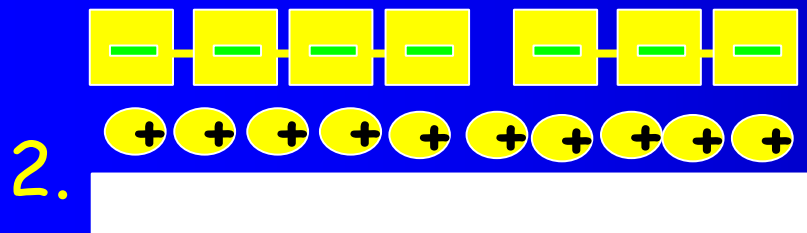


Cuidado com sítios de Poliadenilação!!

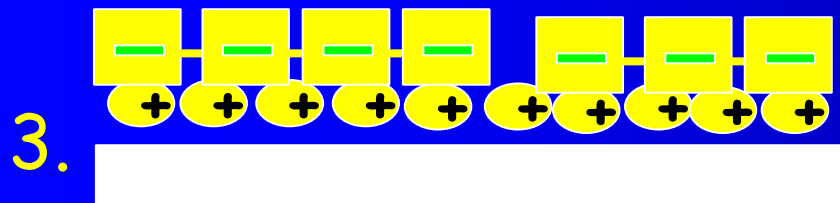
Microarrays: Imobilização dos cDNAs em lâmina de vidro



Lâminas cobertas
com aminosilano



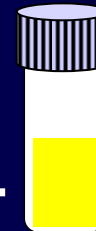
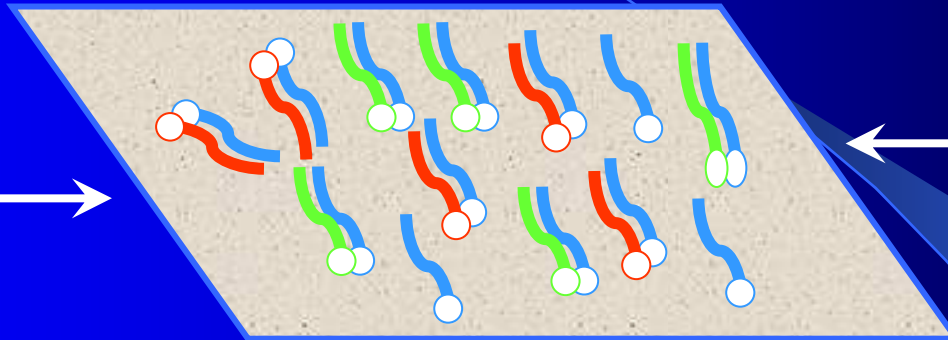
Deposição do DNA



Fixação covalente do DNA
com UV crosslink

Cinética de hibridização

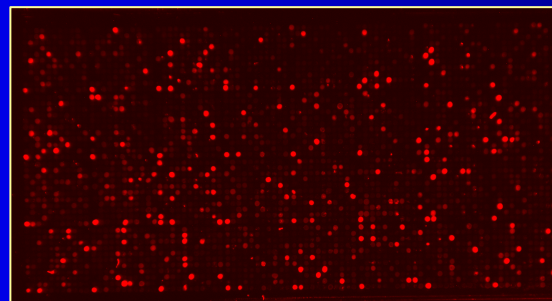
cDNAs



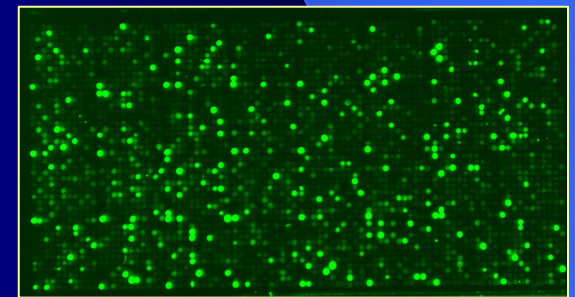
Misturas de cDNAs
marcados *CY3* e *CY5*

Captura das
Imagens

Cy5



Cy3



Medida dos dados

Localização dos *spots*

Quantificação das intensidades
e/ou medidas de qualidade para
cada *spot*

Geração da tabela de dados



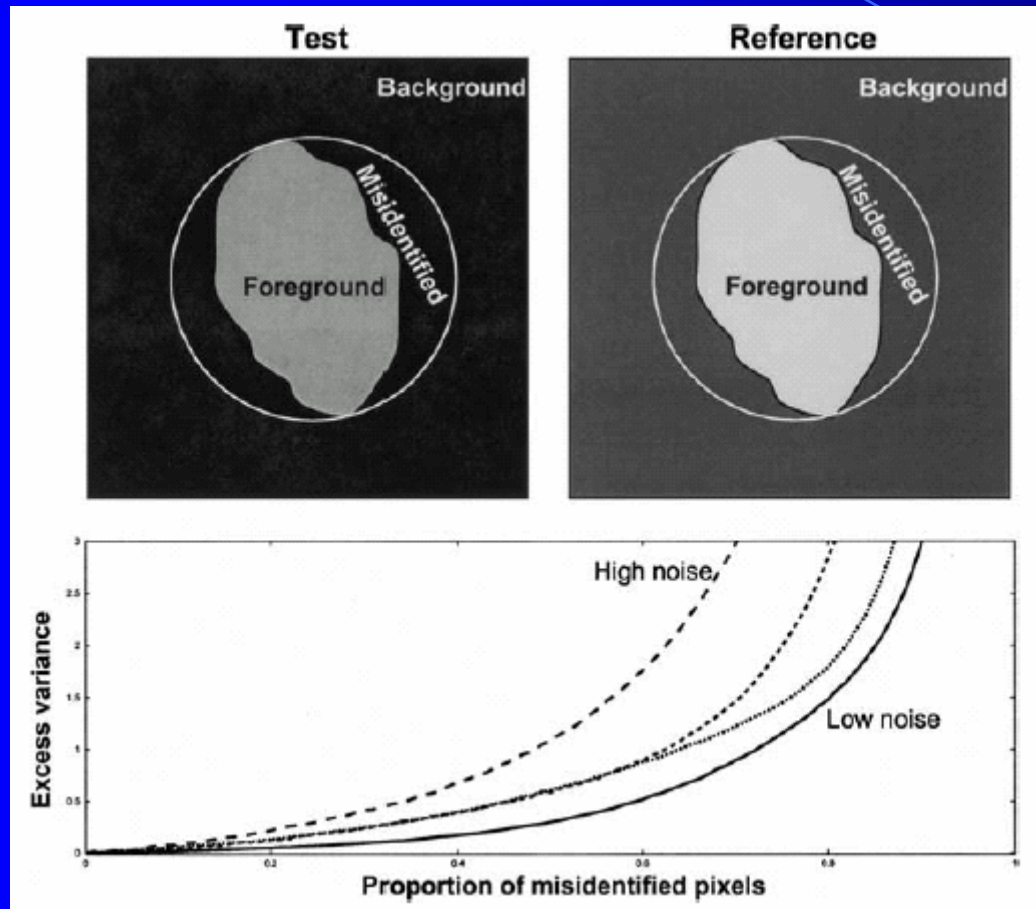
*A medida dos
dados*

Problemas na obtenção dos dados de arrays:

Como identificar os "spots"?

Como escolher os "pixels" para quantificação dos valores de intensidades?

Localização de "spots"



A localização correta dos *spots* leva a resultados mais precisos e reprodutíveis!!

Jain, *et al.* Genome Research, 12: 325-332, 2002

ROIs

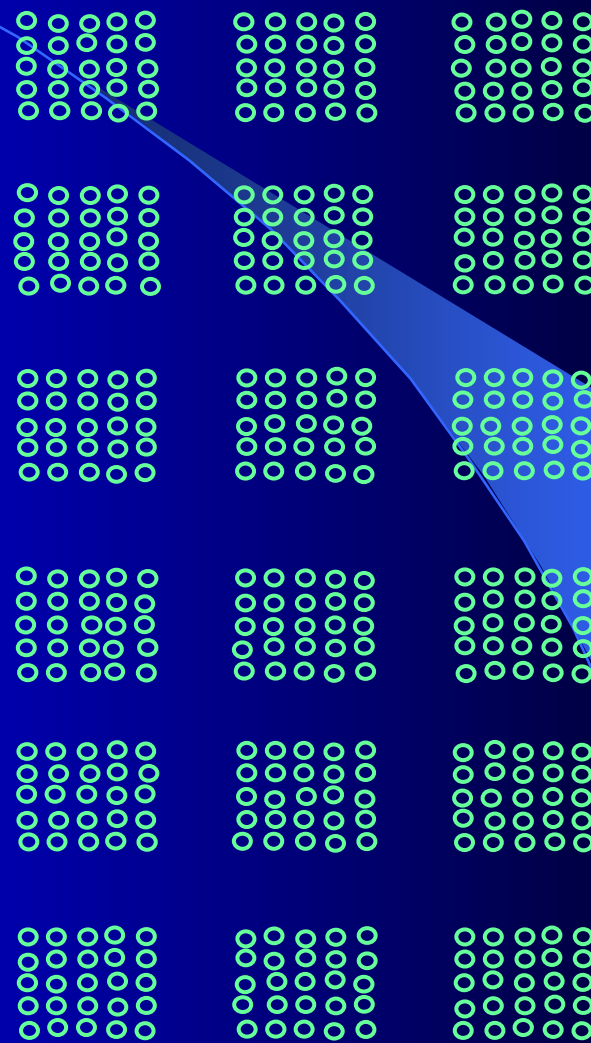
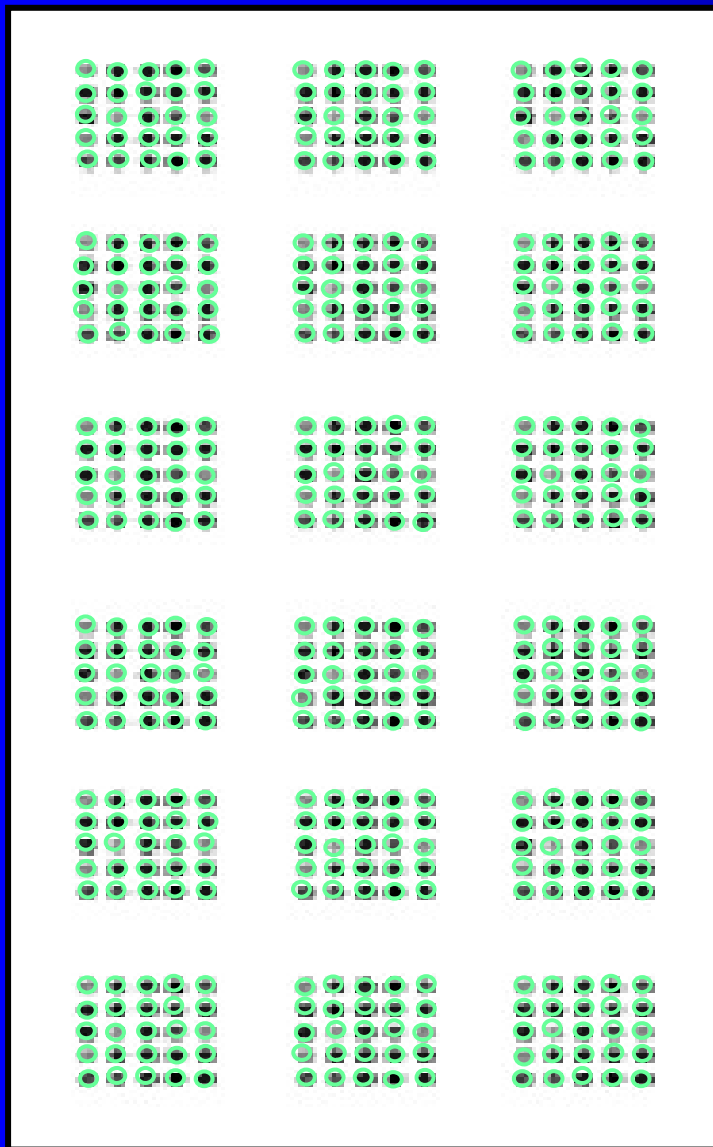
- ✓ *Softwares* comerciais geralmente se baseiam na determinação dos *ROIs* (*Regions Of Interest*).
- ✓ São mais trabalhosos e exigem interferência do usuário (principalmente correção de *spots* mal localizados)



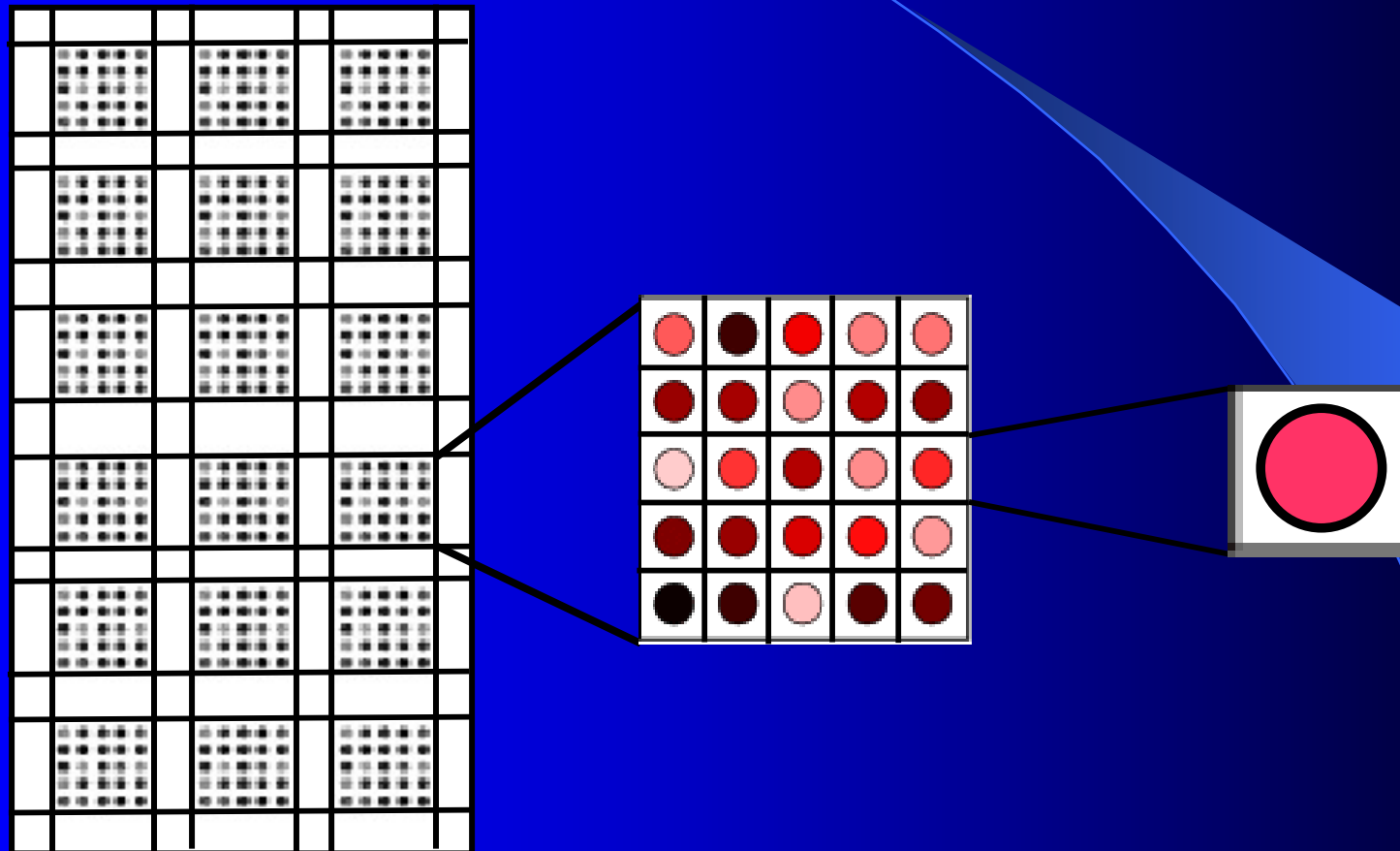
Segmentação

- ✓ A segmentação se baseia em 3 etapas diferentes: localização dos blocos (*subarrays*), localização dos *spots* e detecção da área ocupada por sinal.
- ✓ É um processo automatizado e exige menor interferência do operador
- ✓ Processo 100% reprodutível

Localização de spots: ROIs



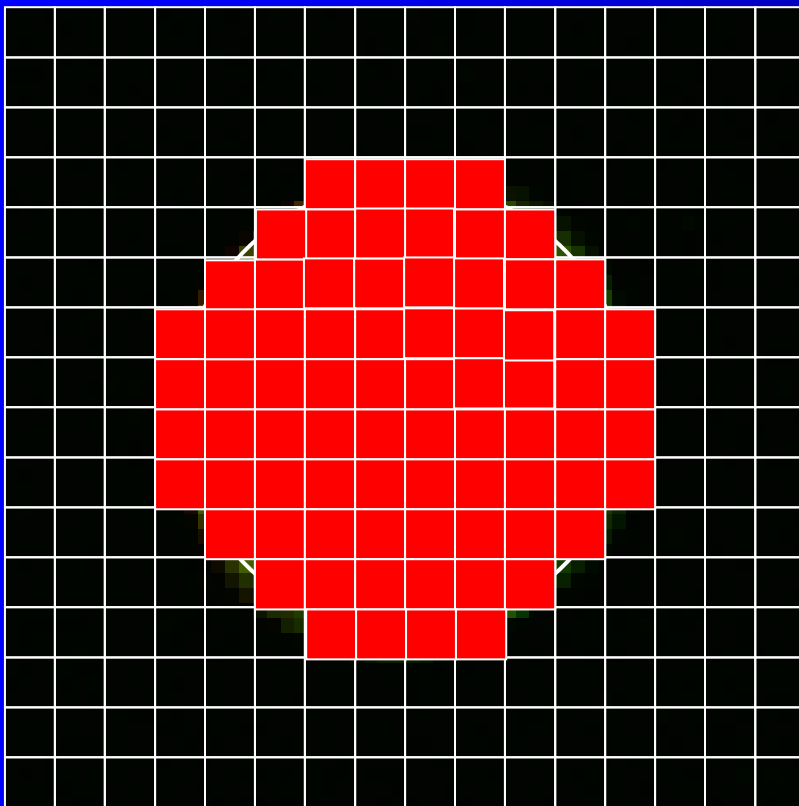
Processo de Segmentação de Imagens de Microarrays



Hirata R, Barrera J, Hashimoto R, Dantas D, Esteves G. *In press*, 2002.

Definição dos "pixels" a serem utilizados para medida de sinal

Uma vez delimitado o *spot* devemos selecionar os *pixels* que serão utilizados para medir os valores de intensidade deste *spot*

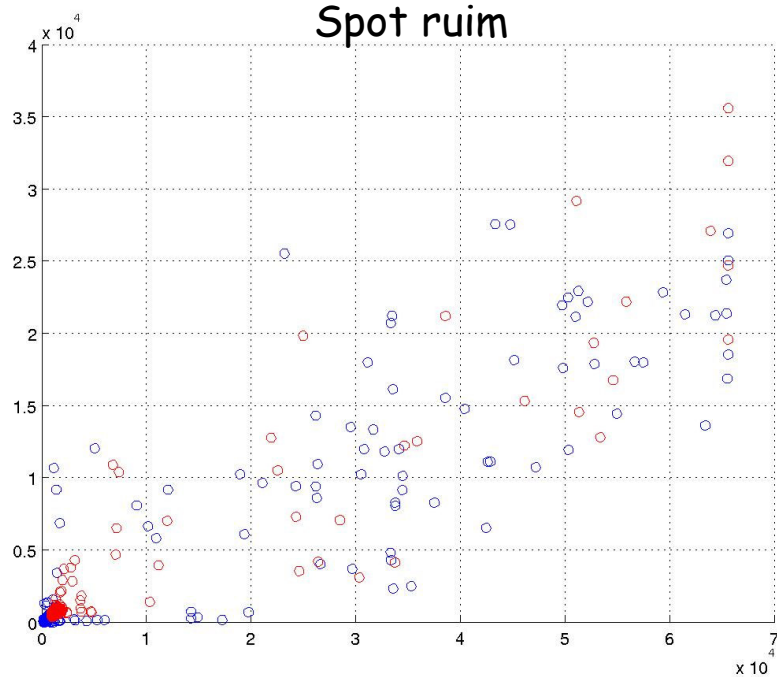


Podemos então selecionar os *pixels* para o *spot* e medir o histograma (distribuições de intensidade)

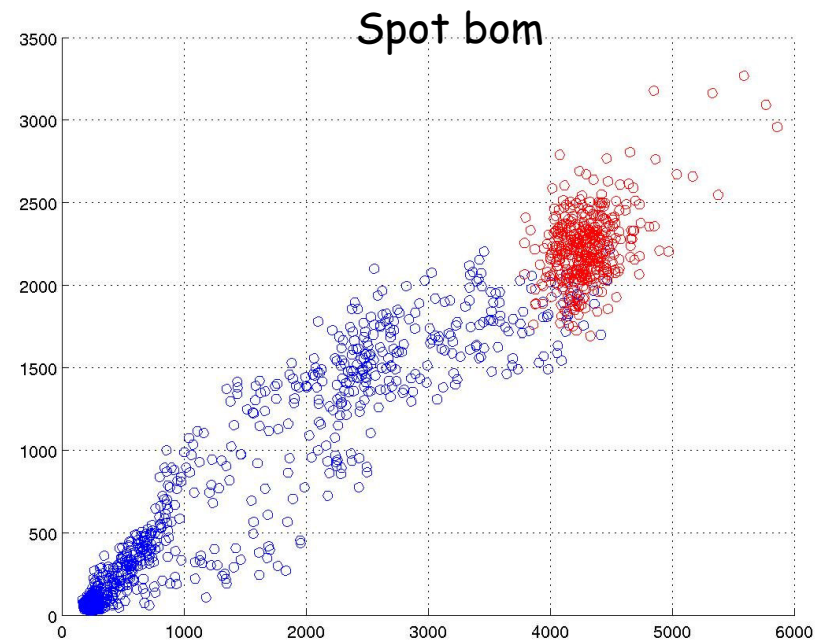
Exemplo:
15% > intensidade do *foreground*

O mesmo acontece para a medida do *background*

A dispersão das intensidades de cy3 e cy5 pode nos dar uma idéia sobre a qualidade dos spots



○ *Pixels do background*
○ *Pixels do foreground*



Expressão diferencial

A razão é a medida mais utilizada para medir a expressão diferencial
Isso pode ser feito de duas maneiras conceitualmente distintas

1.
$$\frac{CH1I - CH1B}{CH2I - CH2B}$$

Medida baseada em *spots*

2.
$$\frac{CH1IP_i - CH1B}{CH2IP_i - CH2B}$$

Medida baseada em *pixels*



Objetivos

Objetivo Geral:

Como objetivo principal, pretendemos comparar os dados obtidos com alguns *softwares* comerciais disponíveis com o software desenvolvido pelo BIOINFO-USP

Objetivos específicos:

- ✓ Comparar a capacidade de definição de áreas
- ✓ Comparar valores de intensidades de sinal e ruído
- ✓ Testar diferentes *softwares* para quantificação de dados de *array*
- ✓ Validação estatística dos dados de *array*



Metodologia

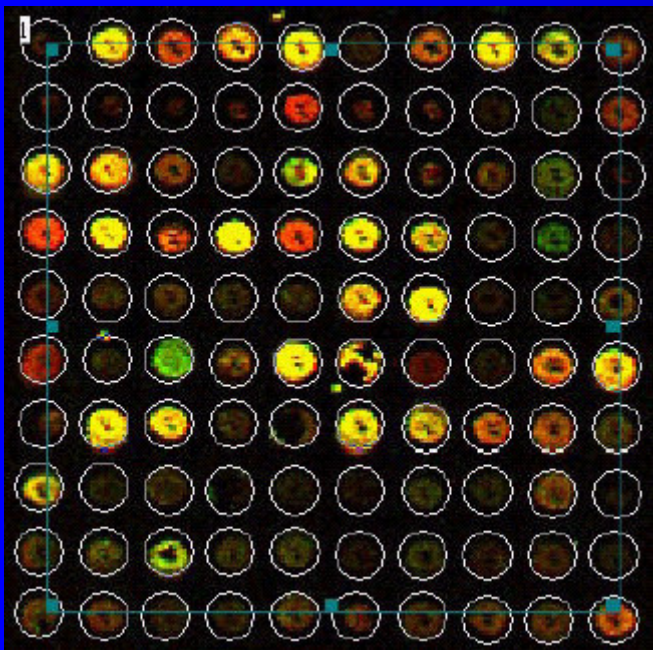
Estratégia Experimental

- ✓ Construir experimentos controlados, onde serão utilizadas quantidades específicas e conhecidas de cDNAs
- ✓ Construir mRNAs sintéticos p/ a hibridização das lâminas
- ✓ Quantificar as imagens obtidas com diferentes *softwares* (scanalyze, quantarray, spot e BIOINFO)
- ✓ Comparar os dados

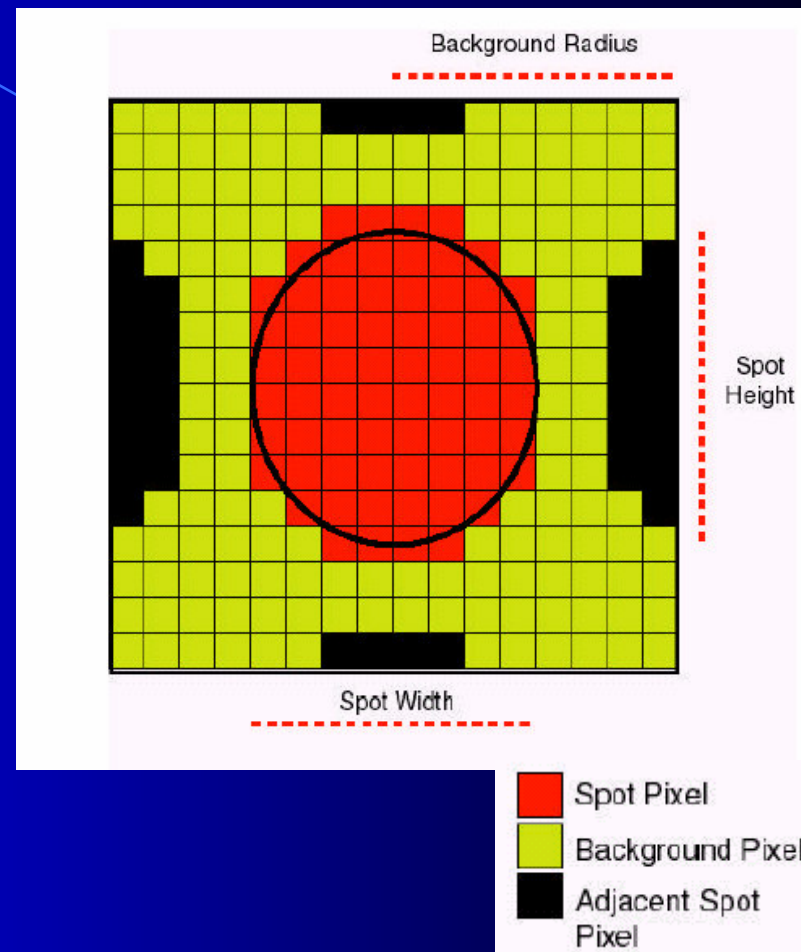
Softwares analisados

- ✓ ScanAlyze 2.44 - Stanford University
- ✓ QuantArray® 3.0 - Packard BioScience
- ✓ Spot 2.0 - UCSF Cancer Center
- ✓ BIOINFO - USP

ScanAlyze



Totalmente baseado em ROIs.



Amostragens de *pixels* simples

ScanAlyze

- ✓ Intensidades (CH1I e CH2I) calculada pela média dos *pixels*
- ✓ *Background* pela mediana dos *pixels* (CH1B e CH2B) ou média (CH1AB e CH2AB)
- ✓ Exporta algumas medidas de relação entre os dois canais (RAT2, MRAT, REGR e LFRAT)

QuantArray®

The screenshot displays the QuantArray software interface. The main window, titled "QuantArray - gugu8x8_genepix.pro", features a menu bar (File, Edit, View, Tools, Window, Help) and a toolbar with various icons. On the left, the "Information Window" is divided into two sections: "Experiment List" and "Analysis Step".

Experiment List:

- GHE040_genepix_1st
- GHE039_genepix_2nd
- GHE039_genepix_1st** (selected)
 - GHE039_Genepix_532_n
 - GHE039_Genepix_635_n
- GHE038_genepix_1st

Analysis Step:

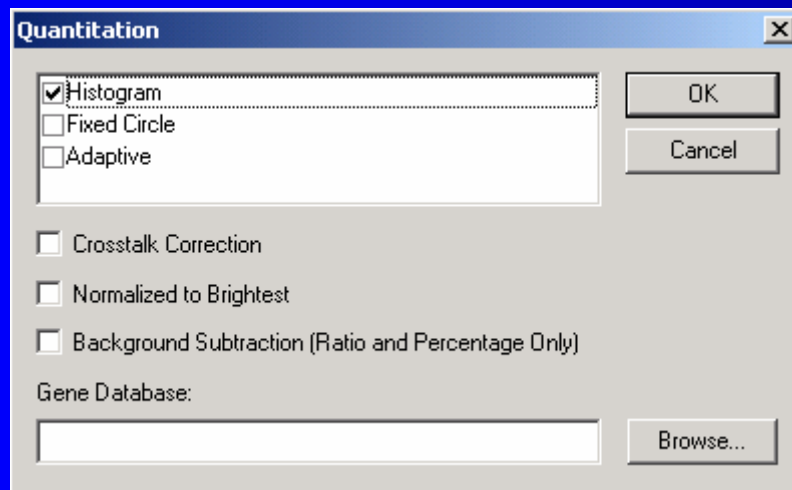
- Register Images
- Specify Location
- Edit Pattern
- Locate Spots** (highlighted)
- View Reports
- Export Data

Below the analysis steps are two buttons: "Start Locate" and "Stop Locate". A checkbox labeled "Use Nominal Locations" is currently unchecked.

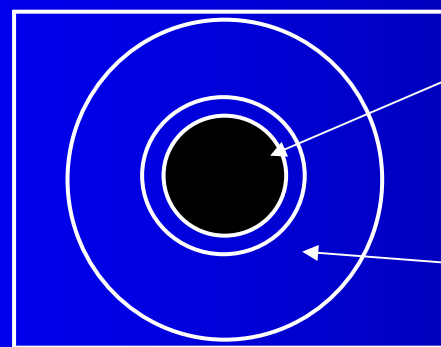
The central area shows a "Composite - 50%" view of a grid of spot images. Each spot is enclosed in a green square frame with a small green crosshair in the center. The spots are arranged in a regular grid pattern across the main window.

At the bottom of the window, the status bar indicates "Shrink to Fit". The Windows taskbar at the very bottom shows the "Iniciar" button, several open applications including "Microsoft PowerPoint - [l]u...", and the "QuantArray - gugu8x8..." window. The system tray on the right shows the date and time as "10:08".

QuantArray® - amostragem de pixels



3 métodos diferentes
de quantificação



Pixels da área de sinal

Pixels da área de background

QuantArray®

Histogram Based Quantitation [X]

Percentile

	Low	High
Signal:	<input type="text" value="80"/>	<input type="text" value="95"/>
Background:	<input type="text" value="5"/>	<input type="text" value="20"/>

Quantification Output

Total Intensities Mode Intensity

Mean Intensity Median Intensity

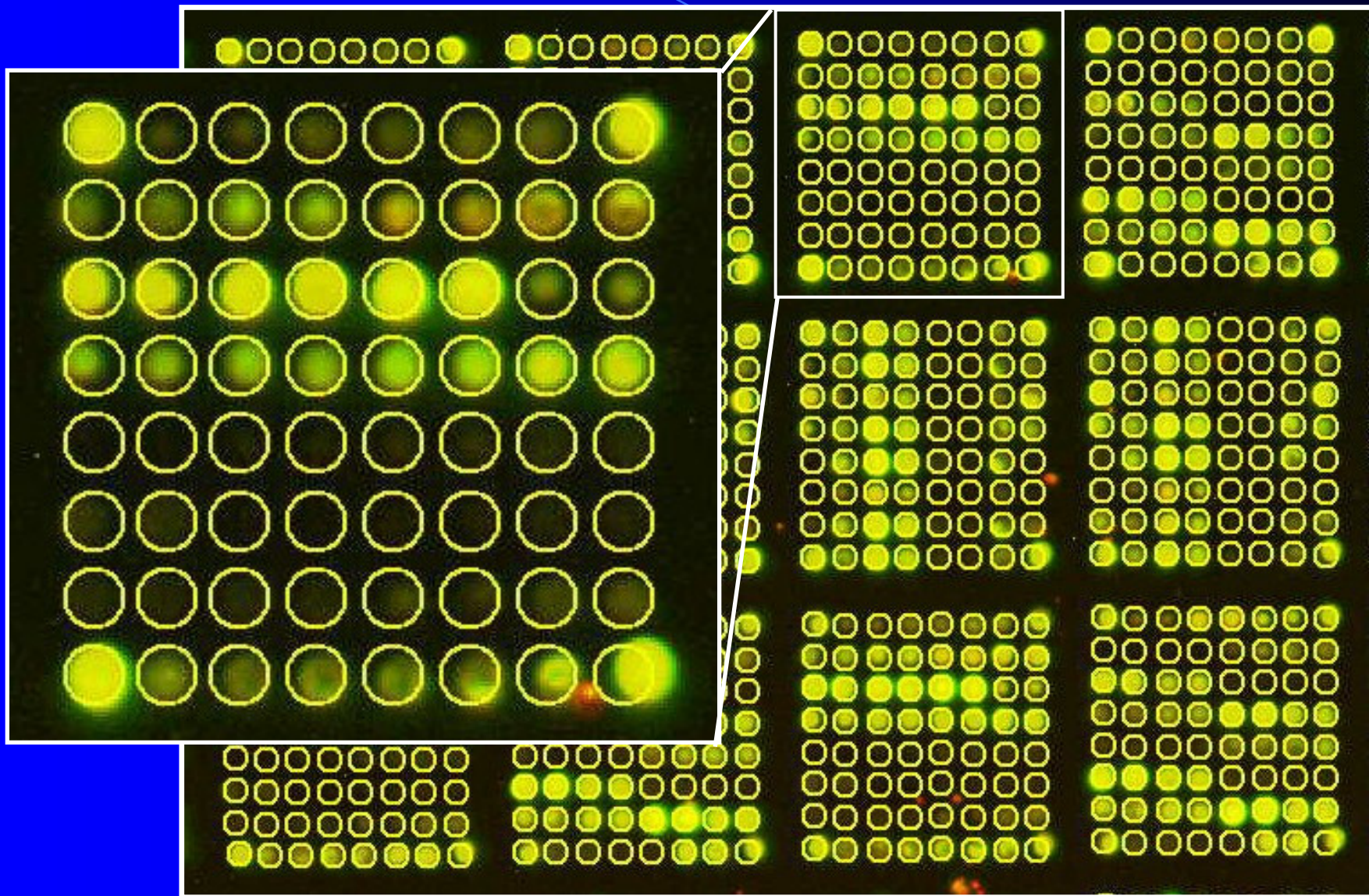
OK

Cancel

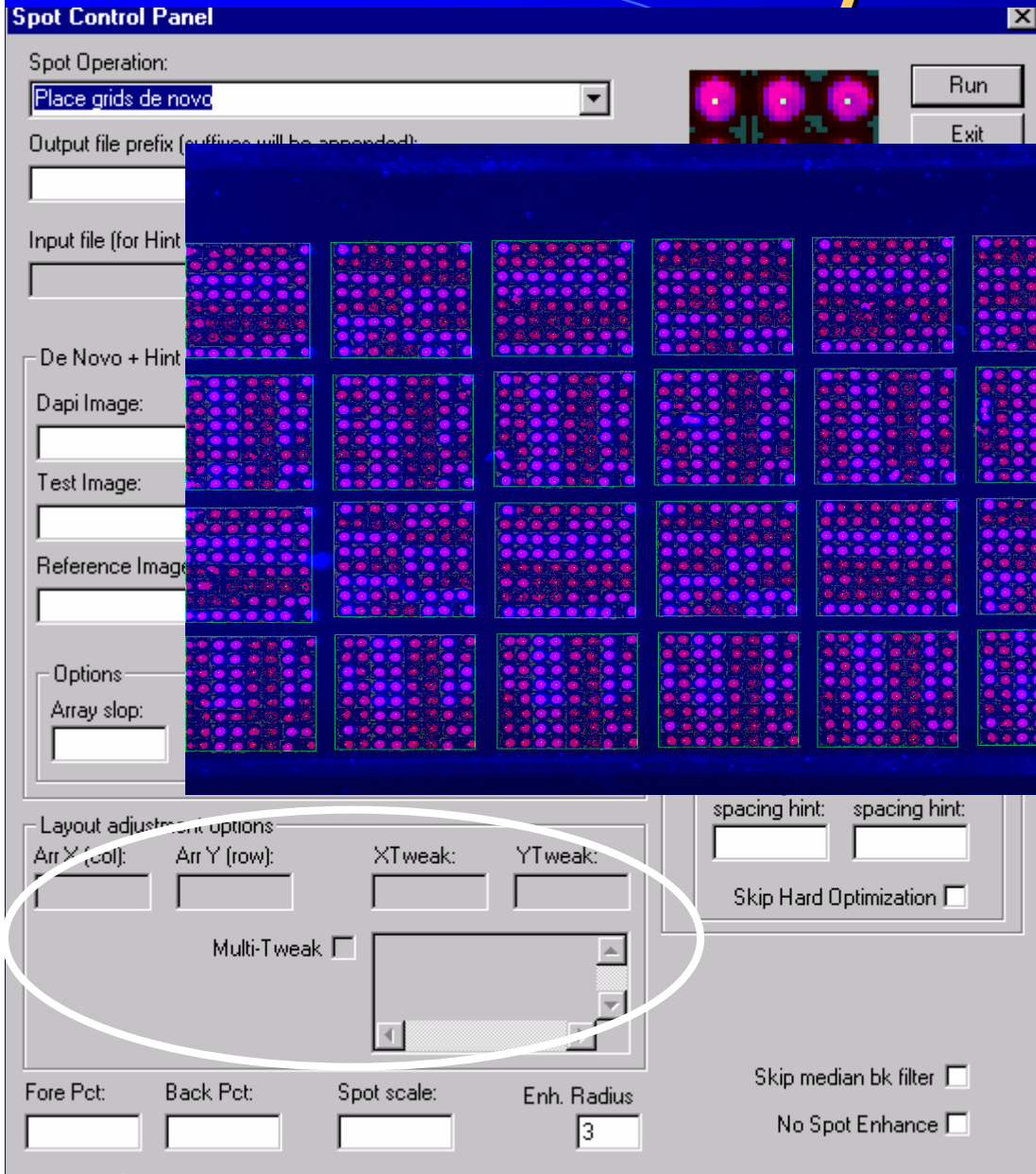
Default

Não exporta nenhuma medida de relação
entre os canais

Exemplo



Spot



Software todo automatizado

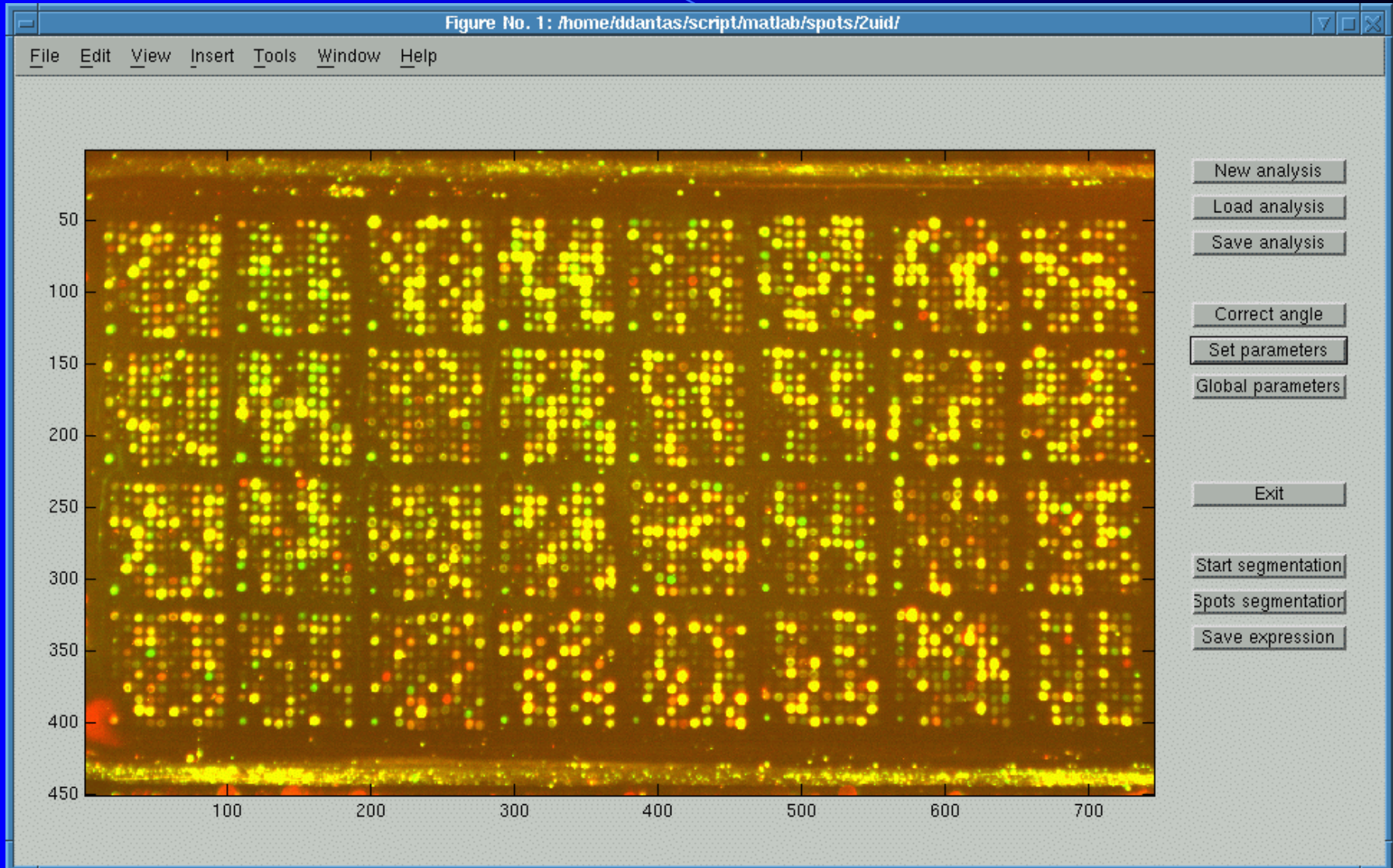
Mas nem sempre funciona...
Então, correção manual!!!

Os pixels são escolhidos com
base no histograma.

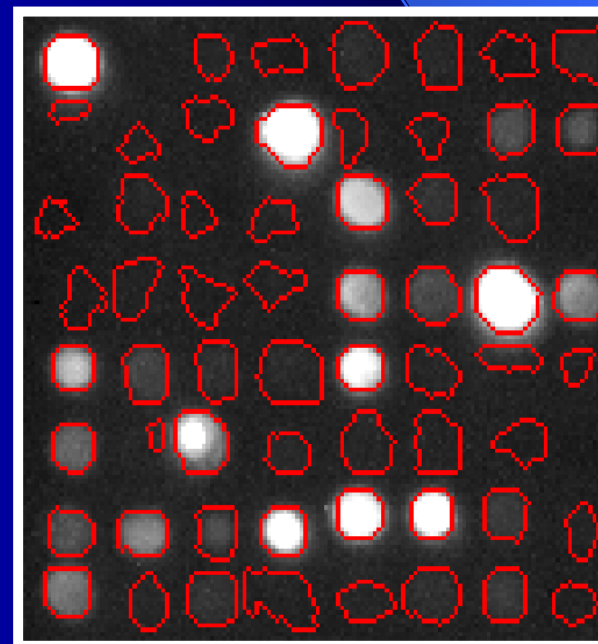
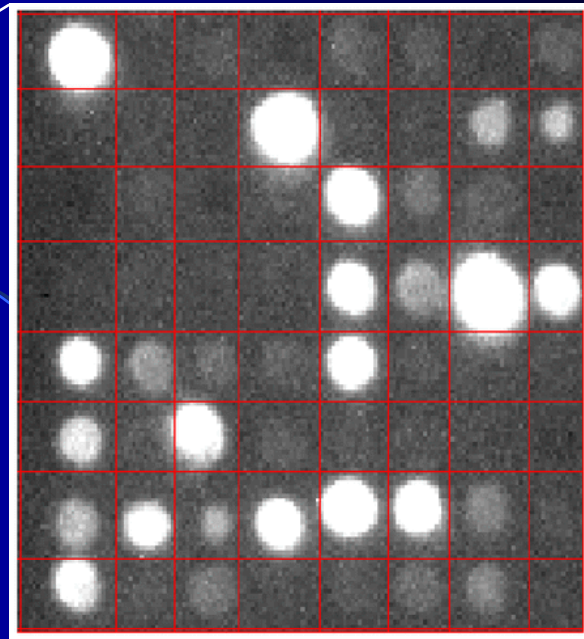
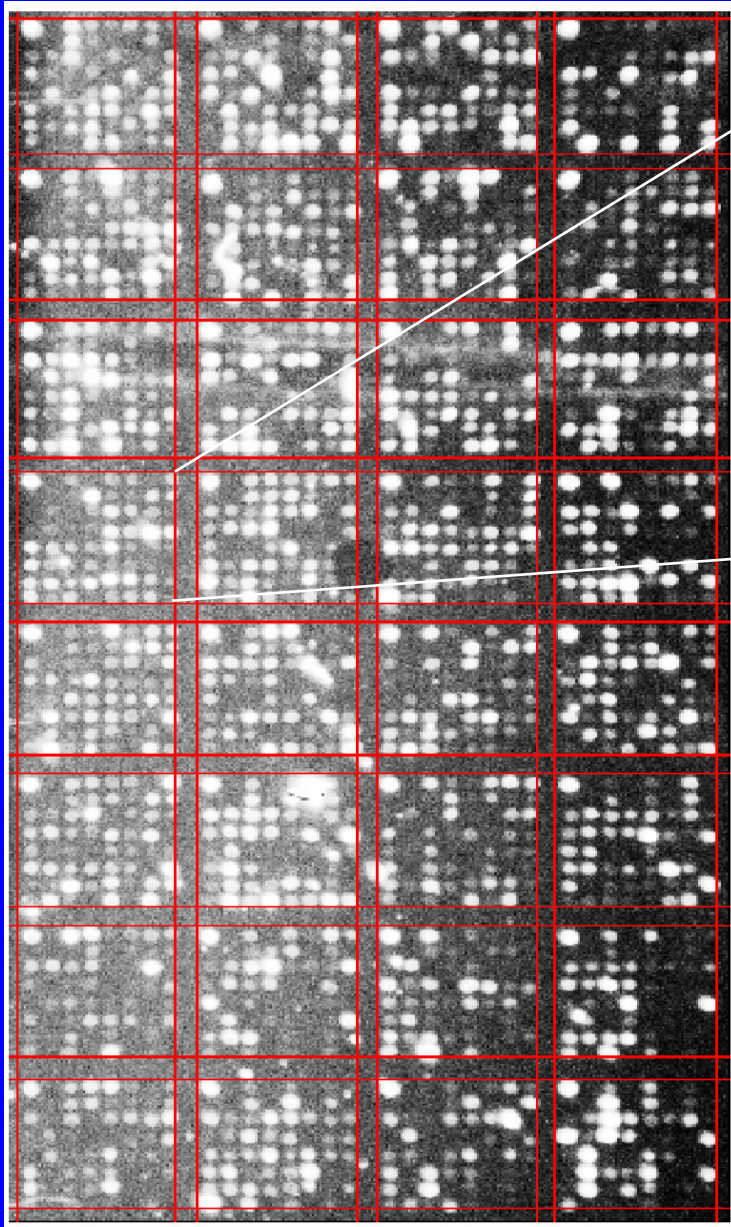
Spot

- ✓ Intensidades calculadas pela média dos *pixels*, tanto para o *foreground* como o *background*
- ✓ Exporta algumas medidas de relação entre os dois canais (RawRat, SpotCorr, Slope, MeanRat, MedianRatio)

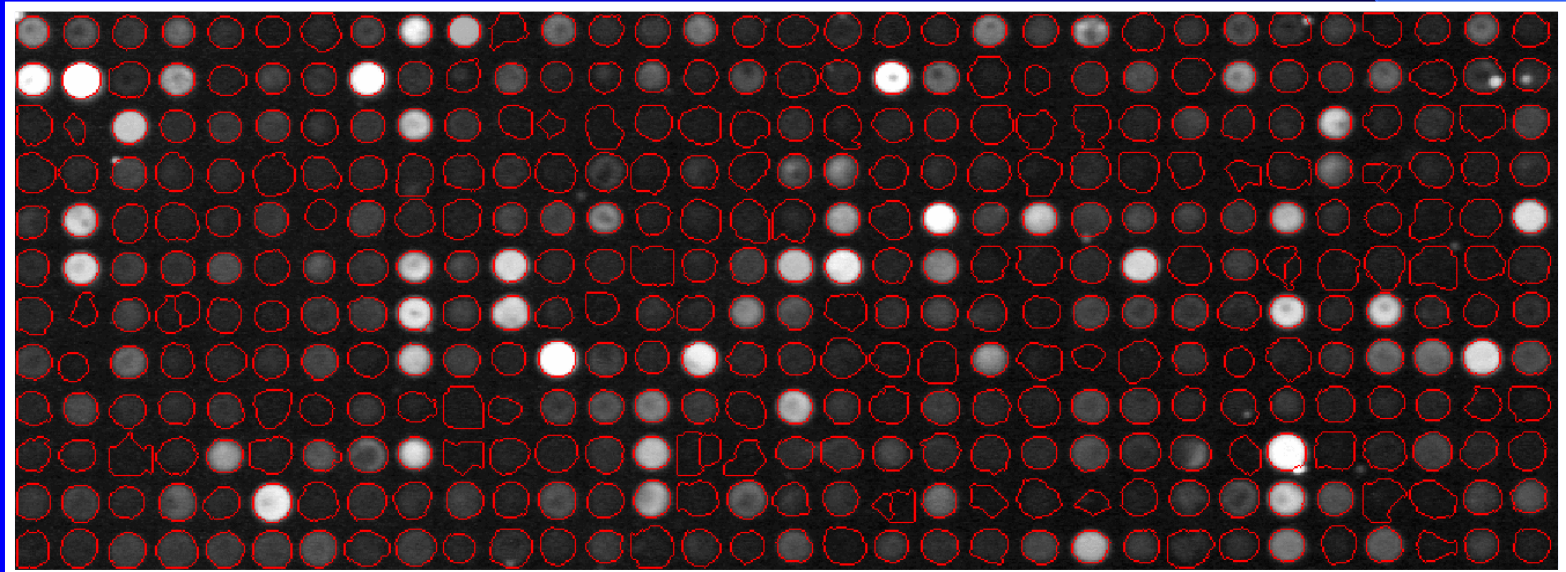
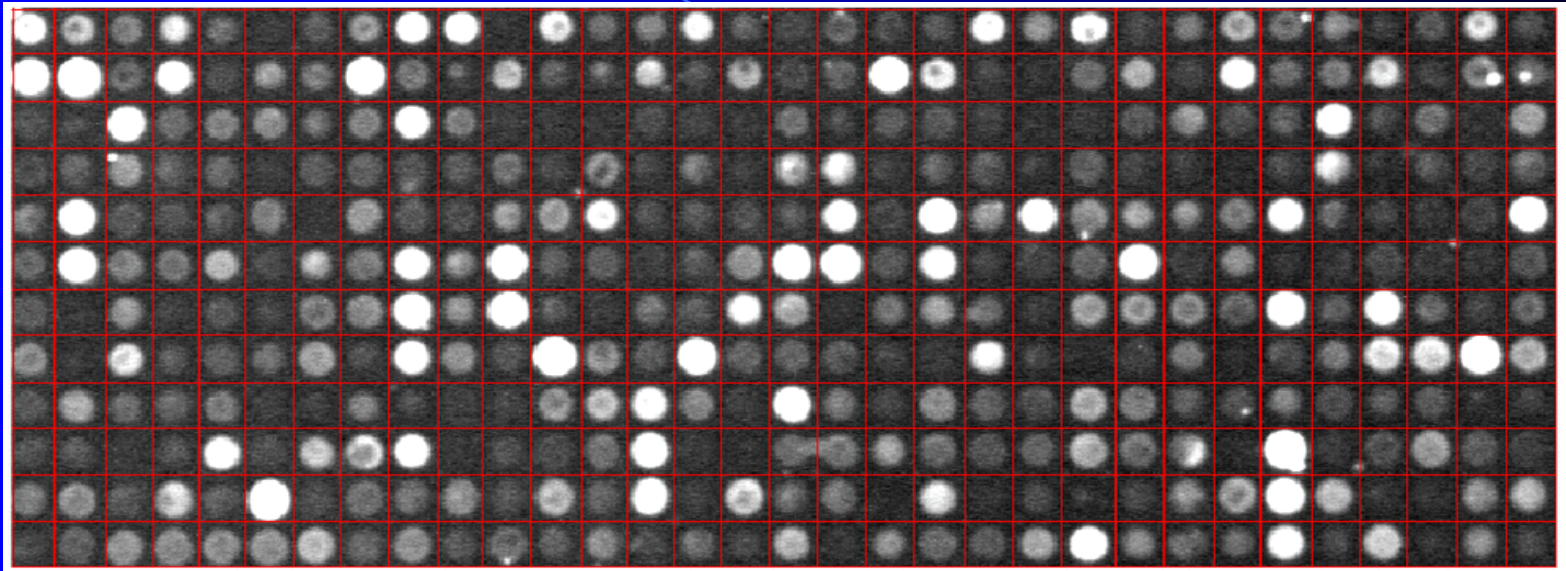
BIOINFO - USP



Exemplo - Segmentação



Exemplo - Segmentação



Bioinfo - USP

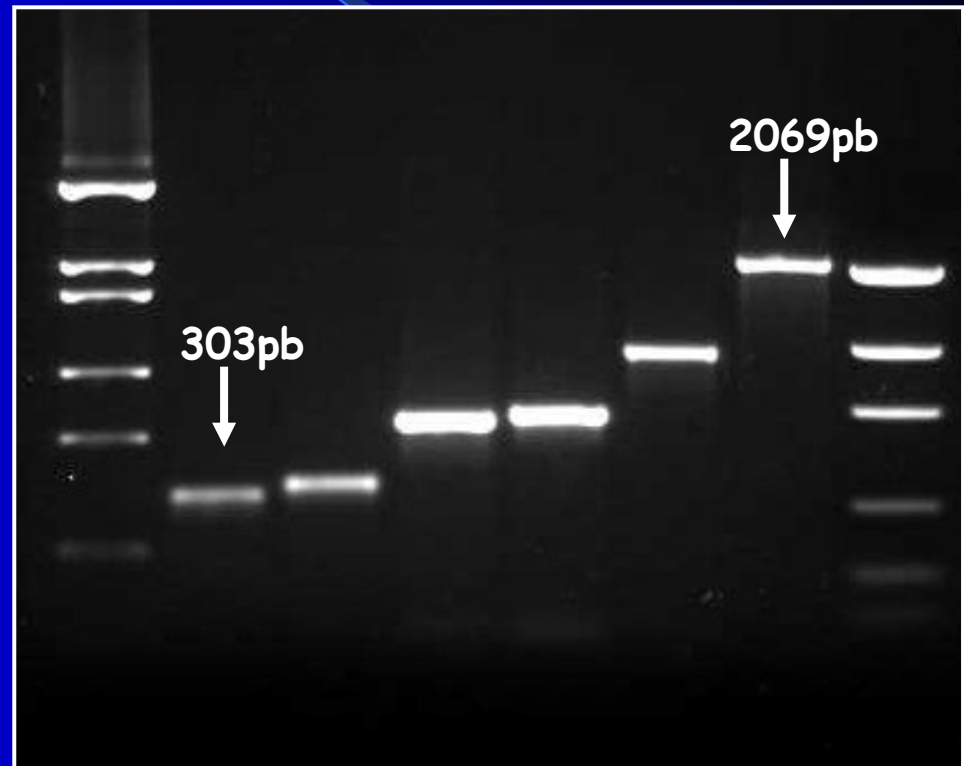
Test name	Best fore	Best back	Standard dev	Mean	Median
GHE037segcy3regmaxccd_maFGuniqueBGnil	-	-	0.12908	-0.036718	-0.021903
GHE037segcy3regmaxccd_maFGsepBGsep_stdMEDiMEDb	96% - 100%	0% - 4%	0.2569	-0.059517	-0.071654
GHE037segcy3regmaxccd_maFGsepBGsep_stdMEDiAVGb	0% - 100%	0% - 4%	0.23874	-0.040205	-0.046444
GHE037segcy3regmaxccd_maFGsepBGsep_stdAVGiAVGb	4% - 100%	0% - 4%	0.23674	-0.030546	-0.041341
GHE037segcy3regmaxccd_maFGsepBGsep_stdAVGiMEDb	4% - 100%	0% - 4%	0.23897	-0.029686	-0.044206
GHE037segcy3regmaxccd_maFGsepBGsep_stdMEDiNILb	0% - 100%	-	0.17139	-0.032841	-0.011528
GHE037segcy3regmaxccd_maFGsepBGsep_stdAVGiNILb	0% - 100%	-	0.1683	-0.024416	-0.0073276
GHE037segcy3regmaxccd_maFGrectBGrect_stdMEDiMEDb	96% - 100%	0% - 4%	0.25481	-0.077784	-0.084916
GHE037segcy3regmaxccd_maFGrectBGrect_stdMEDiAVGb	64% - 100%	0% - 4%	0.19797	-0.063463	-0.061996
GHE037segcy3regmaxccd_maFGrectBGrect_stdAVGiAVGb	4% - 100%	0% - 100%	0.1989	-0.13202	-0.13922
GHE037segcy3regmaxccd_maFGrectBGrect_stdAVGiMEDb	4% - 100%	0% - 4%	0.20197	-0.1319	-0.14125
GHE037segcy3regmaxccd_maFGrectBGrect_stdMEDiNILb	56% - 100%	-	0.12926	-0.030716	-0.013201
GHE037segcy3regmaxccd_maFGrectBGrect_stdAVGiNILb	0% - 100%	-	0.12908	-0.036718	-0.021903
GHE037segcy3regmaxccd_maFGsepcityBGsepcity_stdMEDiMEDb	96% - 100%	0% - 4%	0.31868	-0.060722	-0.10413
GHE037segcy3regmaxccd_maFGsepcityBGsepcity_stdMEDiAVGb	52% - 100%	0% - 4%	0.28825	-0.063455	-0.10329
GHE037segcy3regmaxccd_maFGsepcityBGsepcity_stdAVGiAVGb	4% - 100%	0% - 4%	0.25148	-0.030552	-0.051833
GHE037segcy3regmaxccd_maFGsepcityBGsepcity_stdAVGiMEDb	64% - 100%	0% - 4%	0.28807	-0.065107	-0.10118
GHE037segcy3regmaxccd_maFGsepcityBGsepcity_stdMEDiNILb	36% - 100%	-	0.18429	-0.044763	-0.026774
GHE037segcy3regmaxccd_maFGsepcityBGsepcity_stdAVGiNILb	0% - 100%	-	0.1683	-0.024416	-0.0073276
GHE037segcy3regmaxccd_maFGnilBGnil_stdMEDq	0% - 100%	-	0.17139	-0.032839	-0.011528
GHE037segcy3regmaxccd_maFGnilBGnil_stdAVGq	0% - 100%	-	0.16681	-0.020913	-0.0042654
GHE037segcy3regmaxccd_maFGnilcityBGnil_stdMEDq	36% - 100%	-	0.18499	-0.043333	-0.025813
GHE037segcy3regmaxccd_maFGnilcityBGnil_stdAVGq	0% - 100%	-	0.16727	-0.017649	9.5473e-05

cDNA's Utilizados

- ✓ *lysA* - 303pb, 47,2% de C e G
- ✓ *trpC* - 338pb, 45,3% de C e G
- ✓ Q gene - 637pb, 52,3% de C e G
- ✓ ST0280 (ORESTES) - 659pb, 34,6% de C e G
- ✓ IL6 - 948pb, 37,7% de C e G
- ✓ IRF1 - 2069pb, 52,5% de C e G

Fixação nas Lâminas:

- ✓ Amplificação desses 6 cDNAs por PCR
- ✓ Normalização de suas respectivas concentrações (concentração utilizada $\approx 2,5 \times 10^{13}$ moléculas/ml)
- ✓ 5 diluições para cada cDNA
- ✓ Fixação em posições específicas



Diluições do material fixado

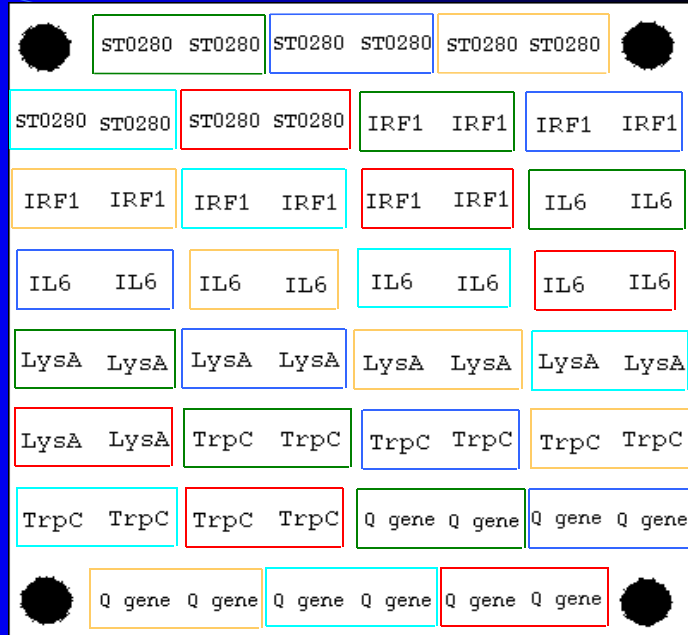
$$d_n(g_i) = \frac{1}{2^{n-1}}, \quad n=1, \dots, 5 \text{ e } i=1, \dots, 6$$

Traduzindo...

Cada cDNA será imobilizado nas diluições de
1/1, 1/2, 1/4, 1/8, 1/16

Desenho dos arrays

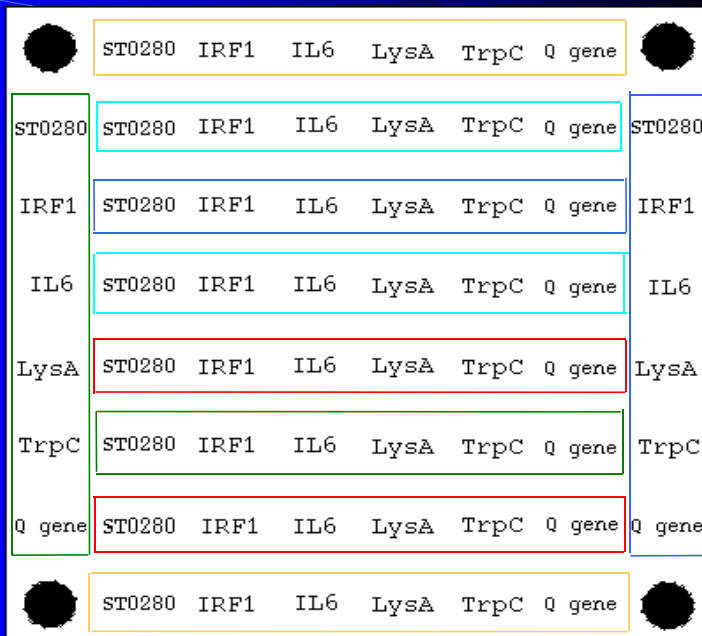
A seguir será mostrado um esquema de como os cDNA's serão fixados nas lâminas.

B1**Legenda**

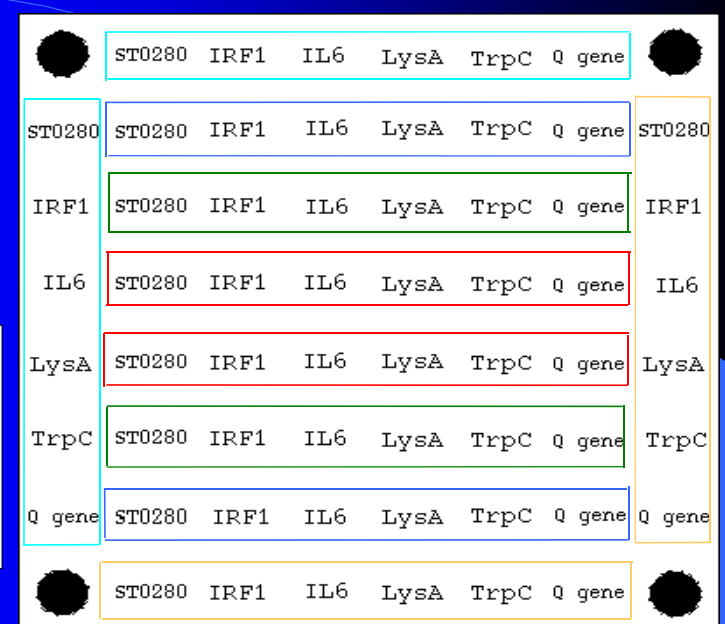
- Diluição 1
- Diluição 2
- Diluição 3
- Diluição 4
- Diluição 5

B2**Legenda**

- Diluição 1
- Diluição 2
- Diluição 3
- Diluição 4
- Diluição 5

B3**Legenda**

- Diluição 1
- Diluição 2
- Diluição 3
- Diluição 4
- Diluição 5

B4**Legenda**

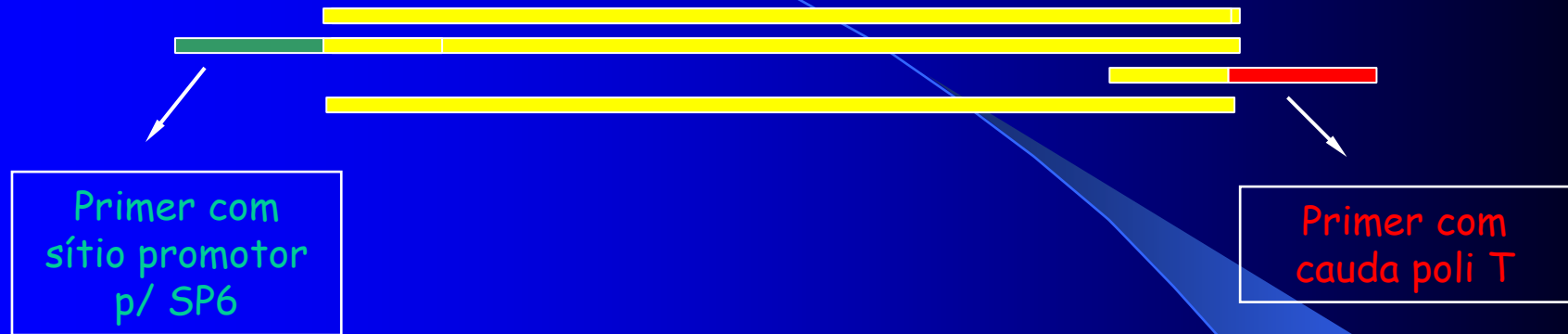
- Diluição 1
- Diluição 2
- Diluição 3
- Diluição 4
- Diluição 5

A 4x8 grid of boxes containing labels B1, B2, B3, and B4. The grid is divided into four quadrants by dashed red lines. The top-left and bottom-right quadrants are 2x4 grids of B1 and B2 boxes. The top-right and bottom-left quadrants are 2x4 grids of B3 and B4 boxes.

B1	B2	B1	B2	B1	B2	B1	B2
B3	B4	B3	B4	B3	B4	B3	B4
B1	B2	B1	B2	B1	B2	B1	B2
B3	B4	B3	B4	B3	B4	B3	B4

Construção de mRNA sintético

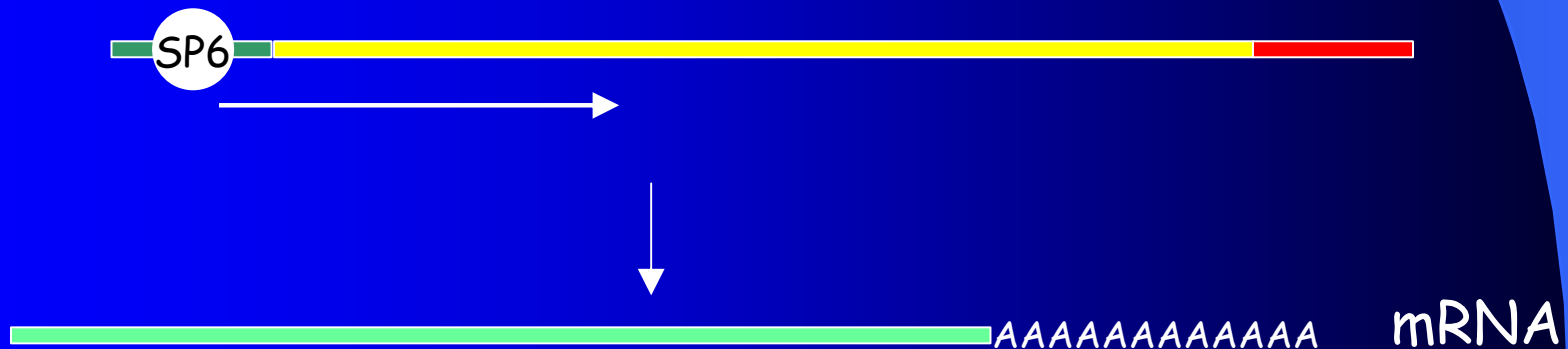
1. Reação de PCR com primers especiais



2. Fragmento de cDNA amplificado com sítio promotor na extremidade 3' e cauda poli T na outra extremidade

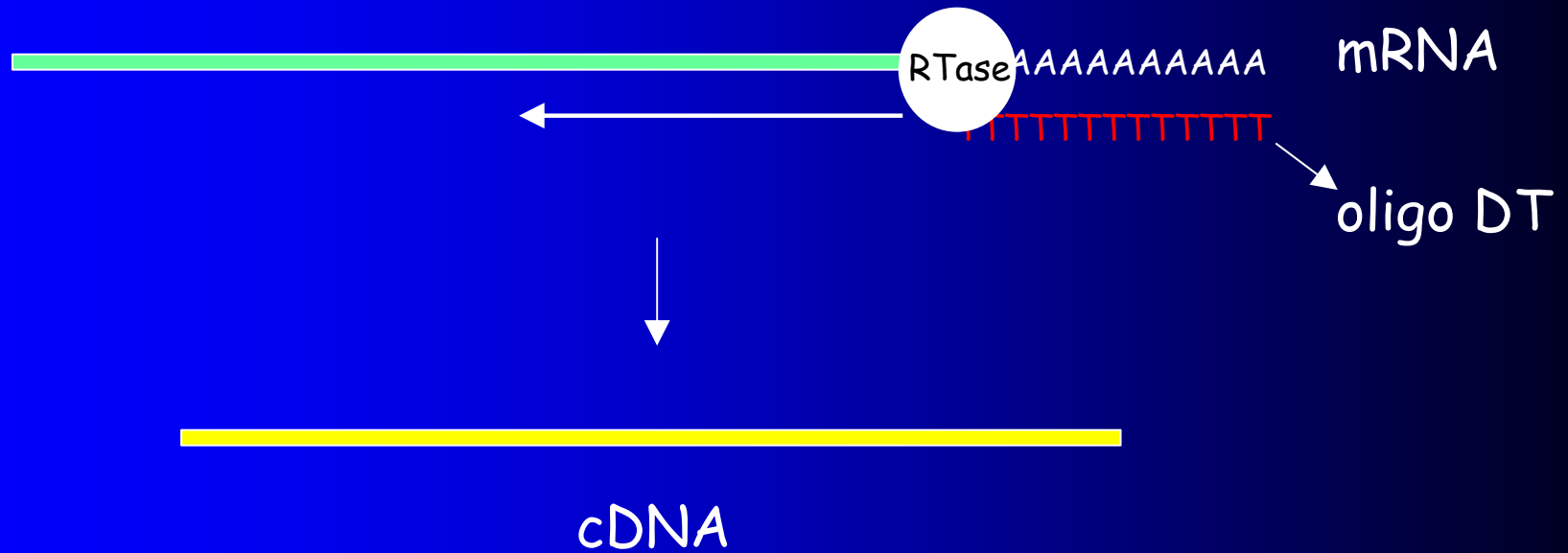


3. Obtenção do mRNA a partir de transcrição *in vitro*



Síntese de cDNA para hibridização

Transcrição reversa



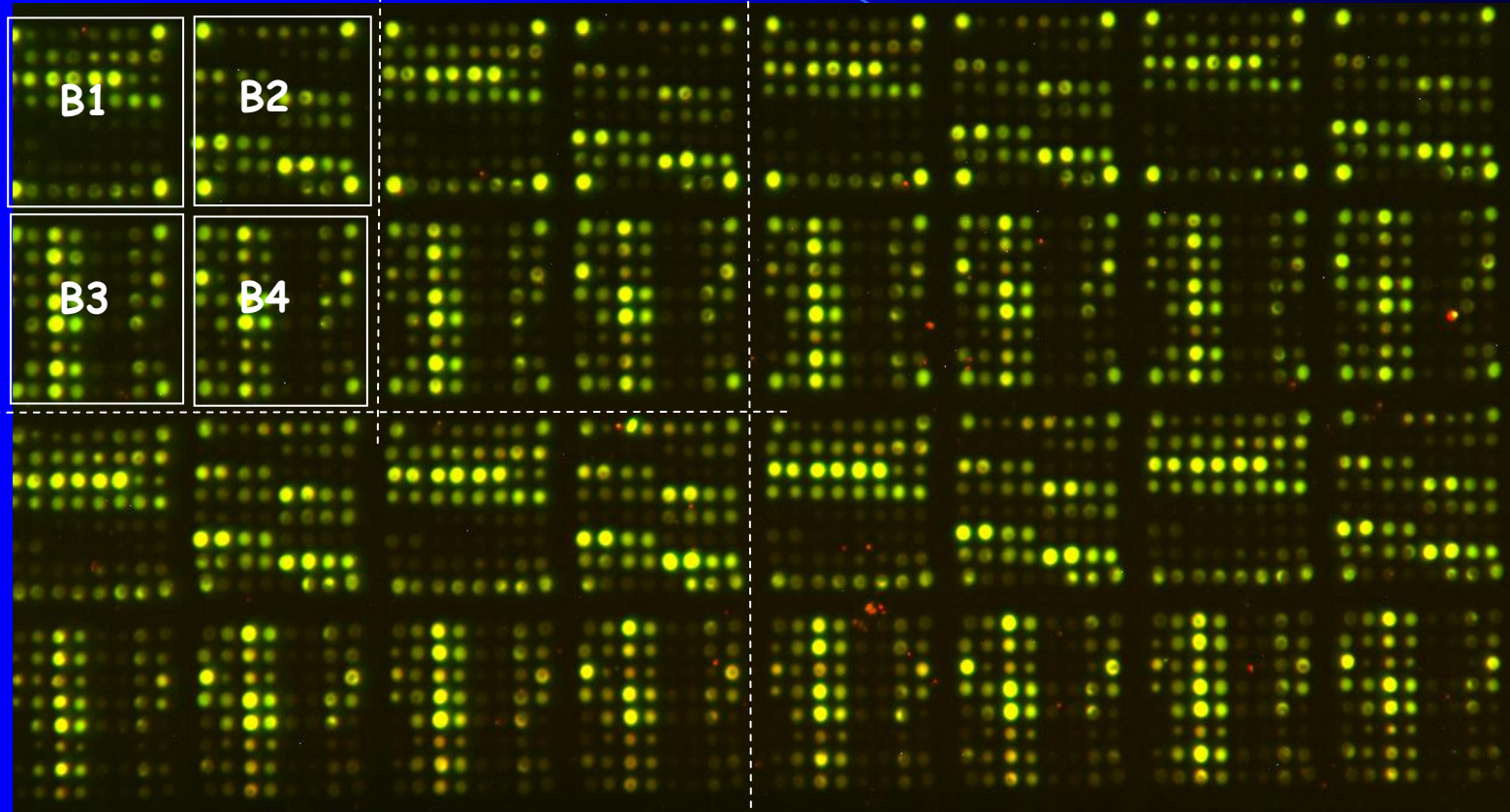
CDNAS marcado

Hibridização



*Resultados
Preliminares*

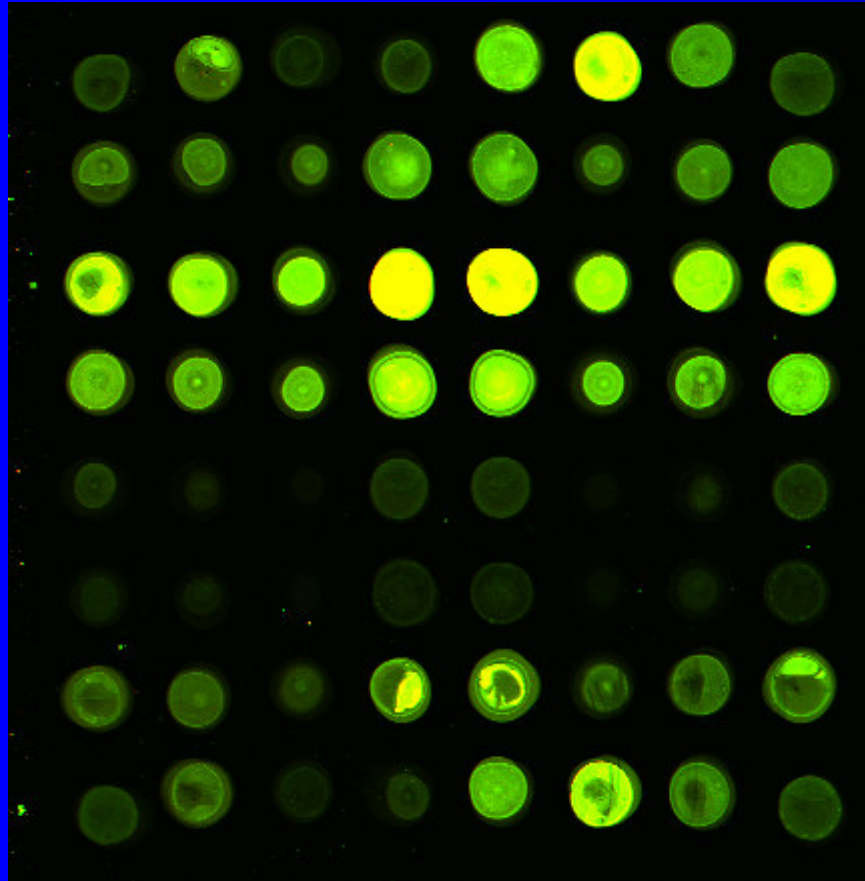
Resultados Preliminares



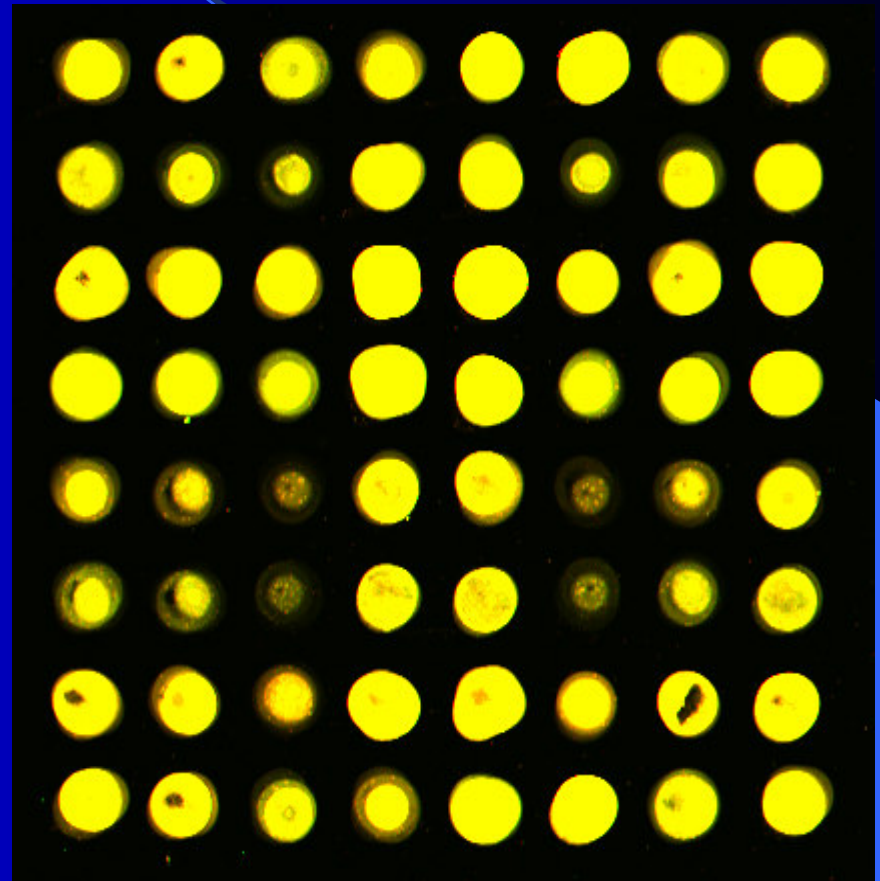
Obs: Quantidade da sonda em excesso comparada com o material fixado

Resultados Preliminares

CCD



Laser - Genepix



Resultados Preliminares

Temos um total de 11 lâminas hibridizadas

- ✓ 2 digitalizadas no CCD (10^{10} e 10^{11} mol. de mRNA)
- ✓ 3 digitalizadas em laser (10^7 , 10^9 e 10^{11} mol. de mRNA)

Cy3 X Cy5
1 : 1

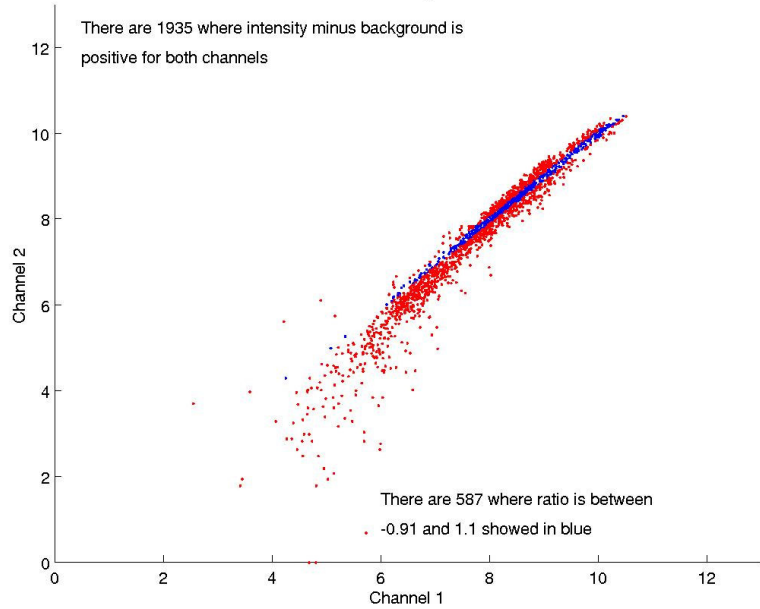
- ✓ 1:3(cy3/cy5) X 1:3(cy5:cy3)
- ✓ 1:6(cy3/cy5) X 1:6(cy5:cy3)

Digitalizadas em laser
 10^9 mol. de mRNA

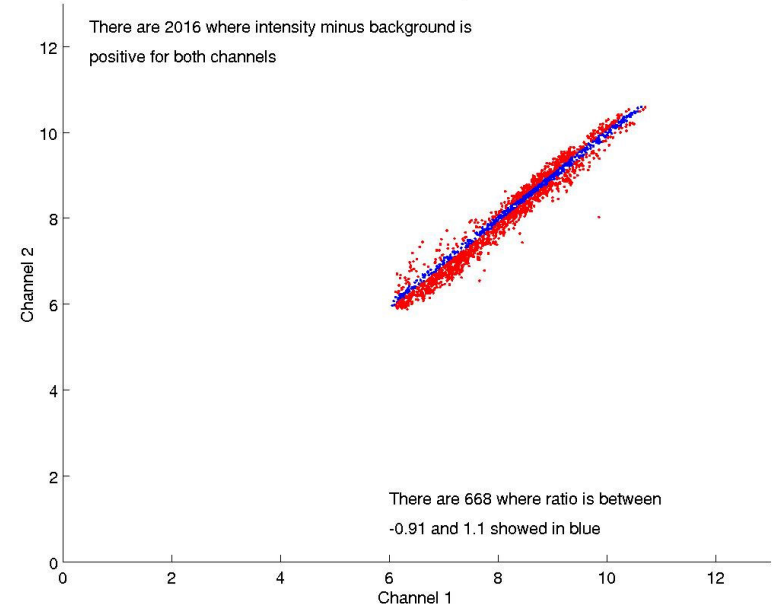
- ✓ Fragmentos comparáveis na proporção de 1:1 e 1:5, com "swap"

Digitalizadas em laser
 10^9 mol. de mRNA

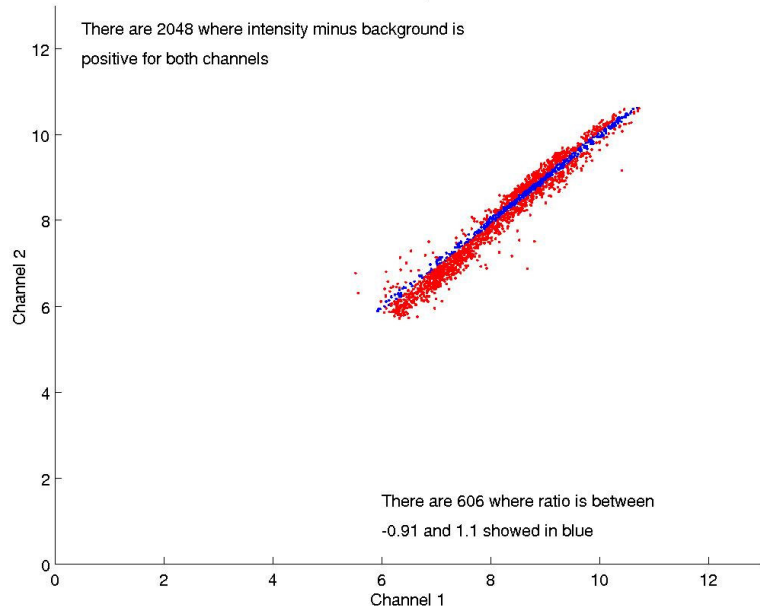
GHE037 2 CCD Scanalyze med corrected



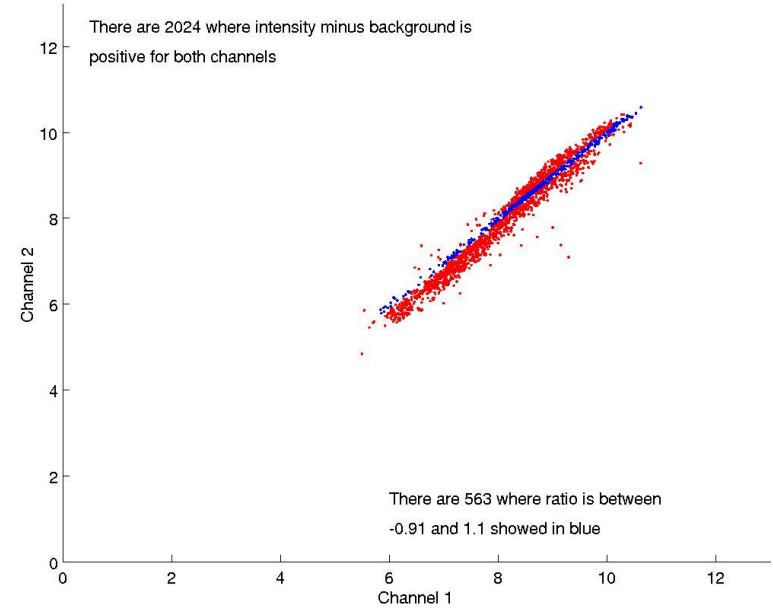
GHE037 2 CCD Quantarrayhist corrected

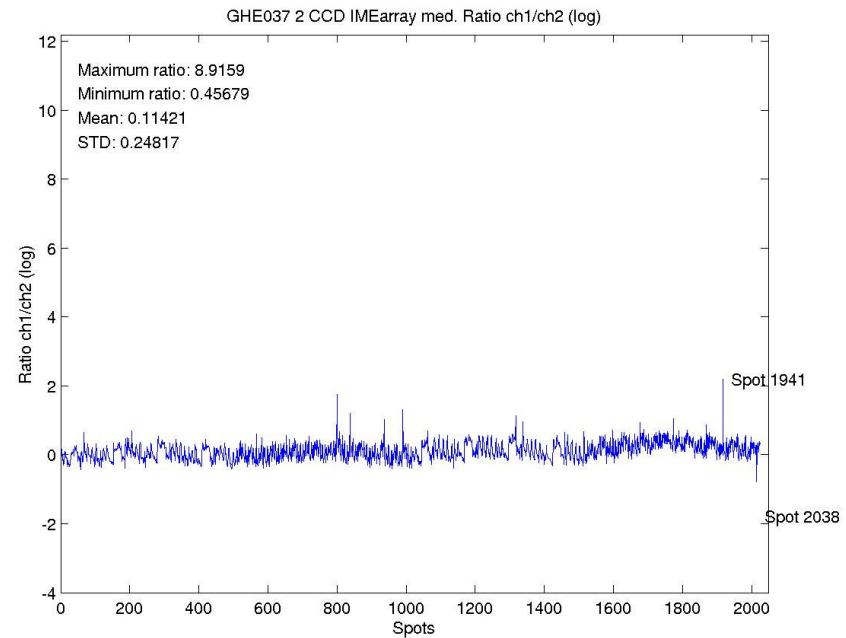
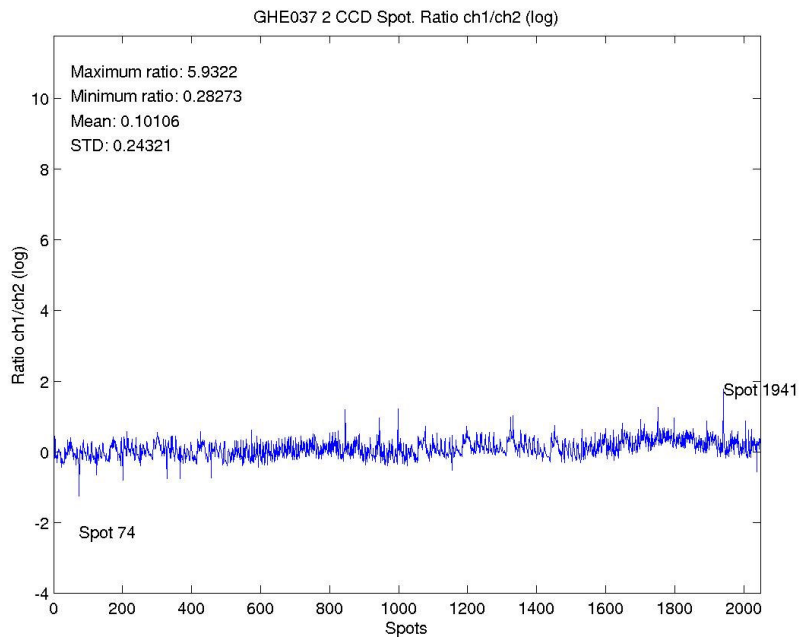
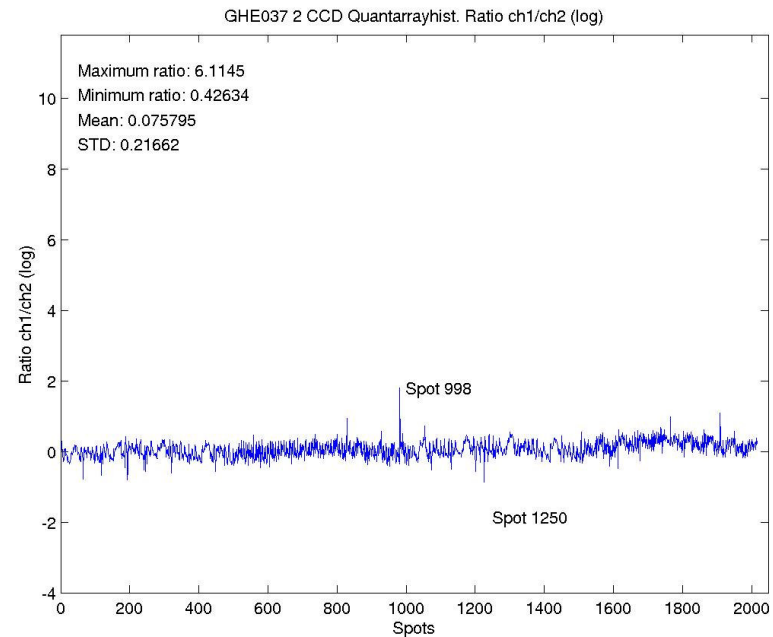
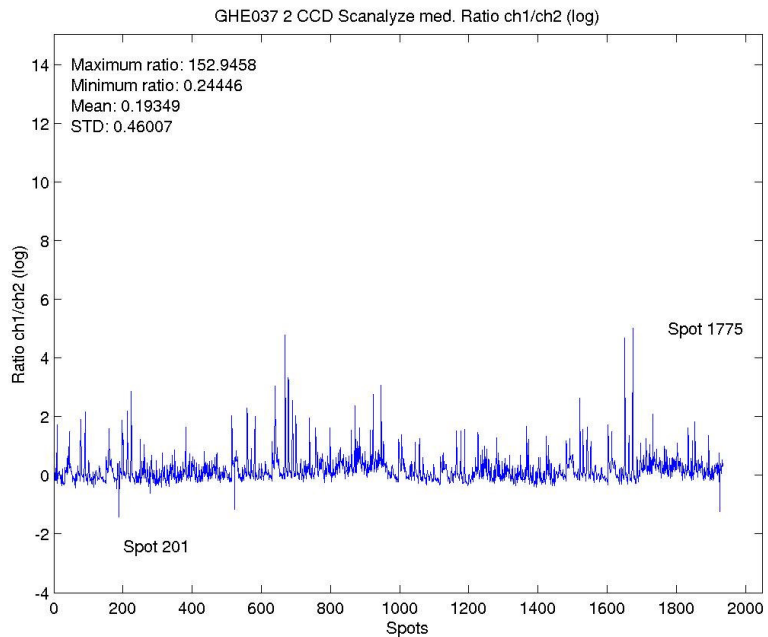


GHE037 2 CCD Spot corrected

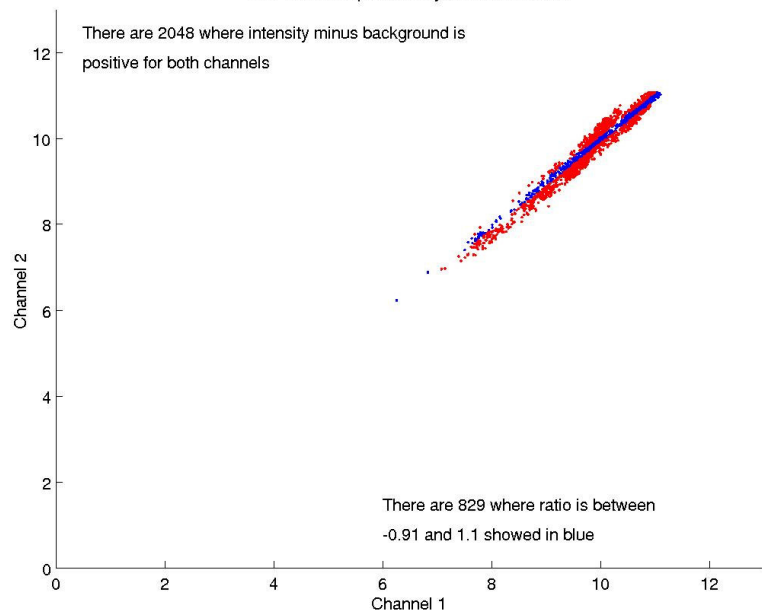


GHE037 2 CCD IMEarray med corrected

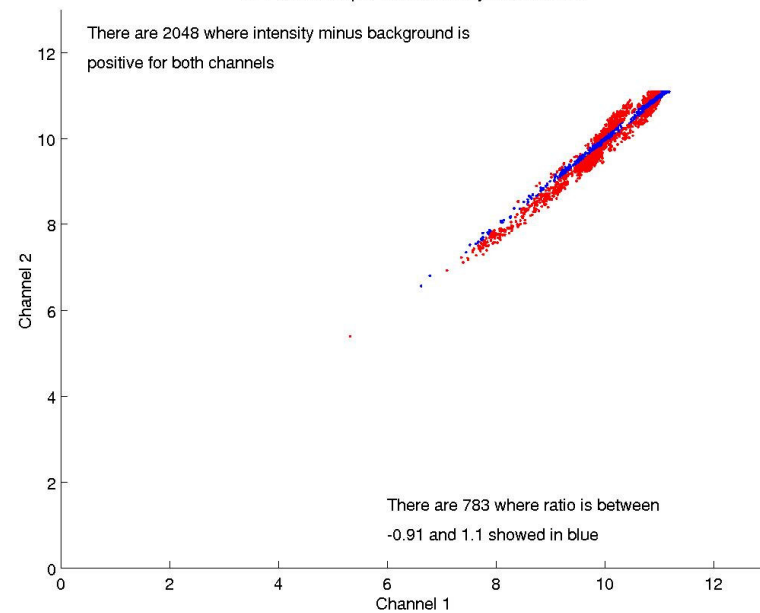




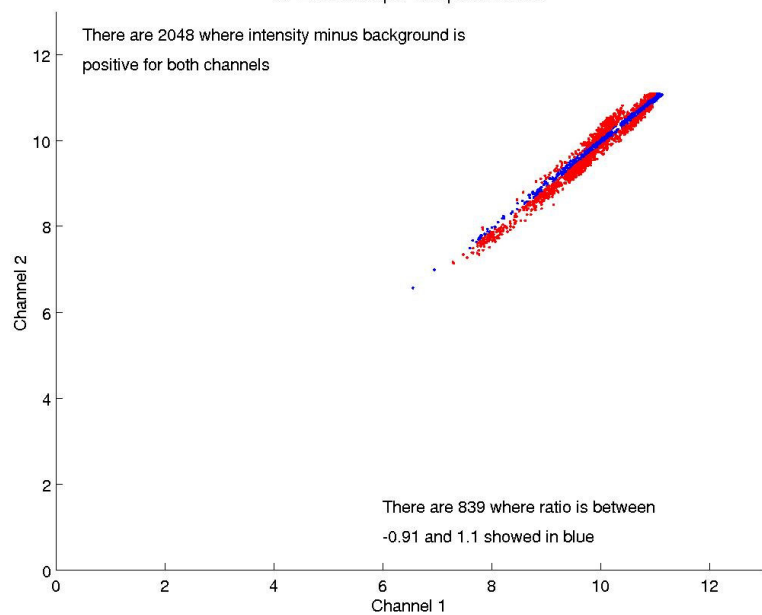
GHE038 Genepix Scanalyze med corrected



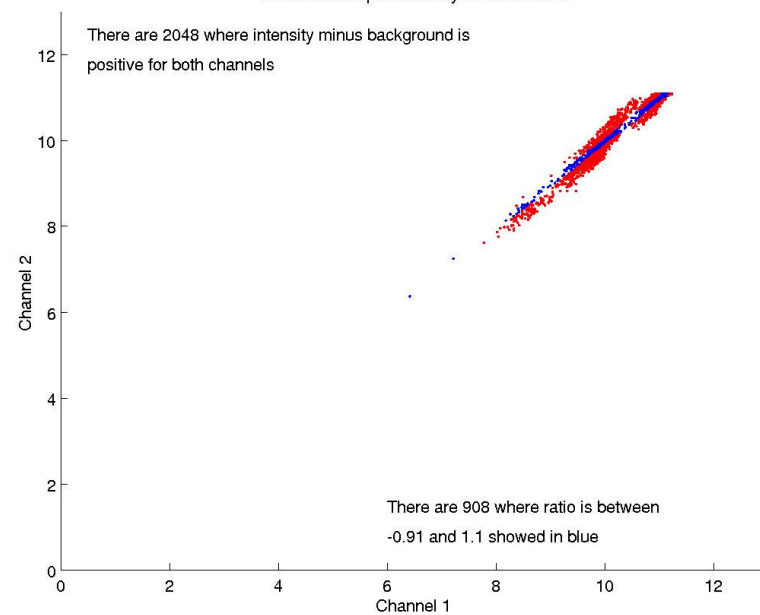
GHE038 Genepix 1st Quantarrayhist corrected

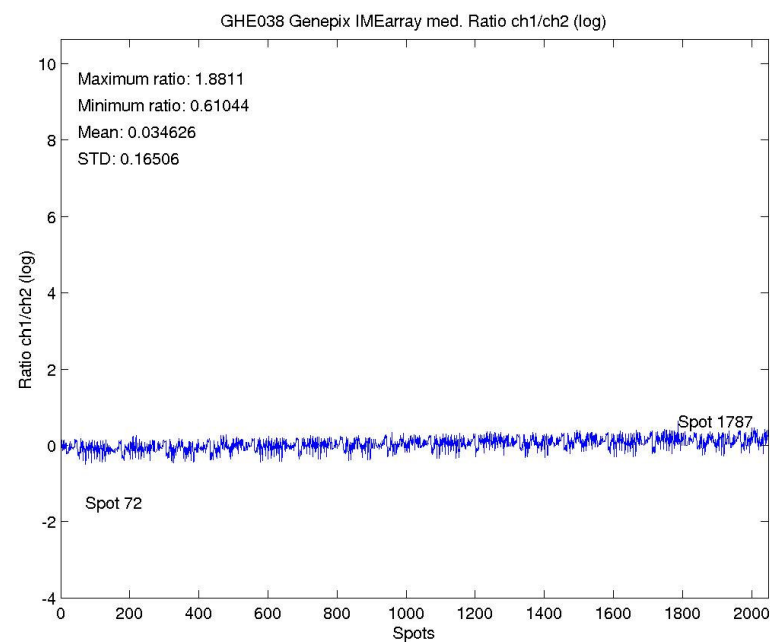
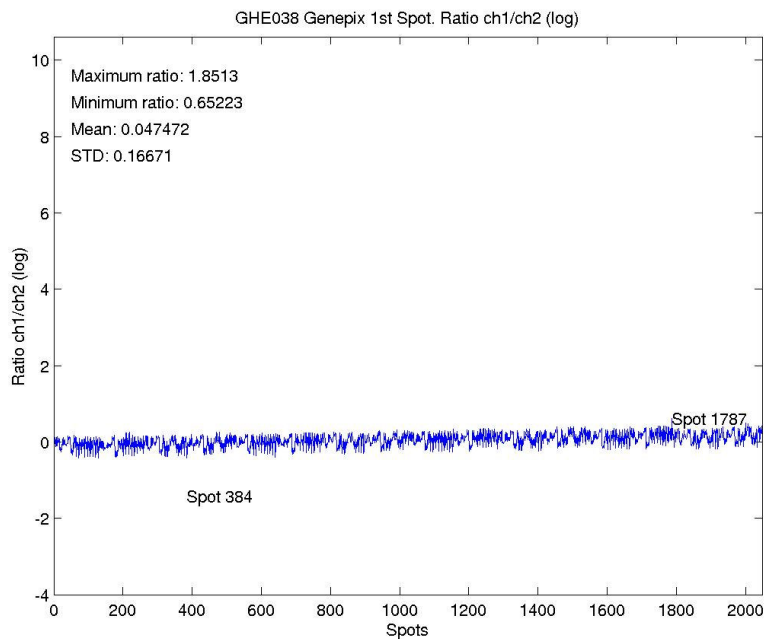
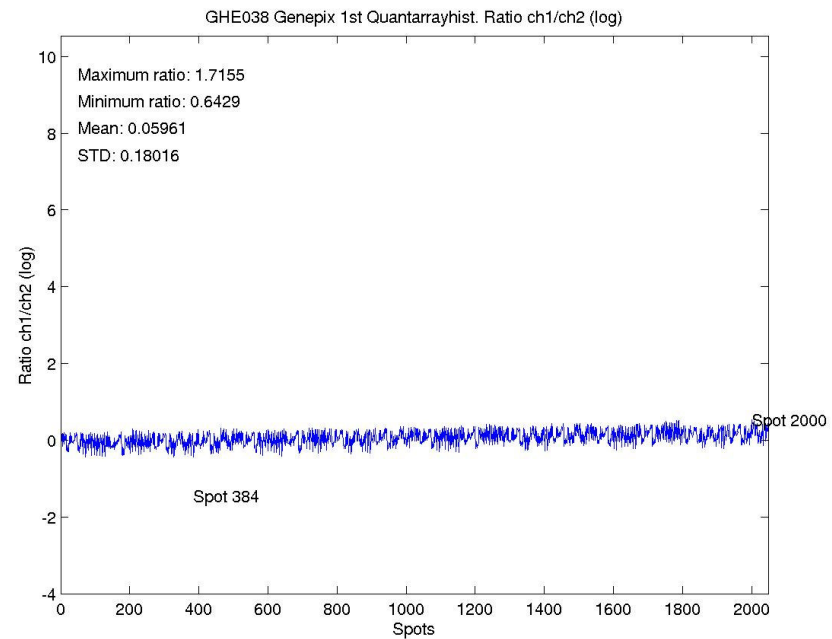
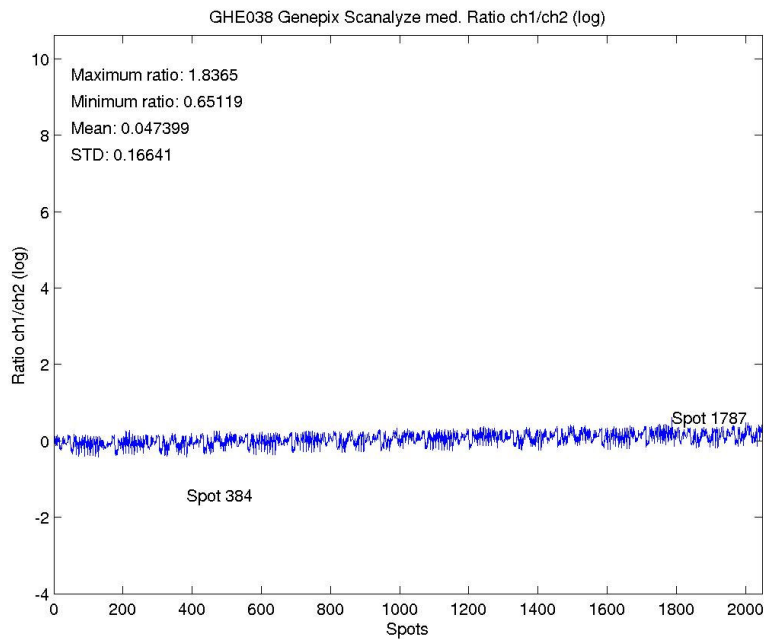


GHE038 Genepix 1st Spot corrected

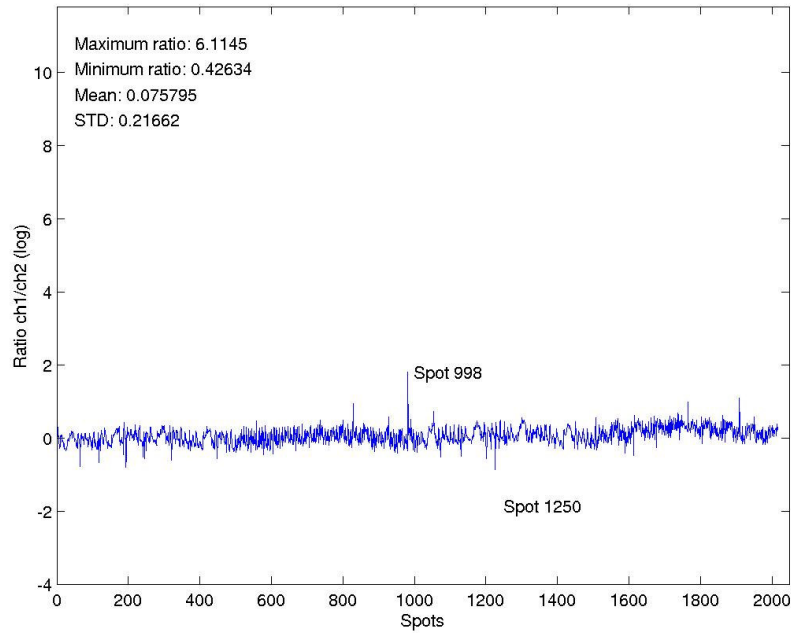


GHE038 Genepix IMEarray med corrected

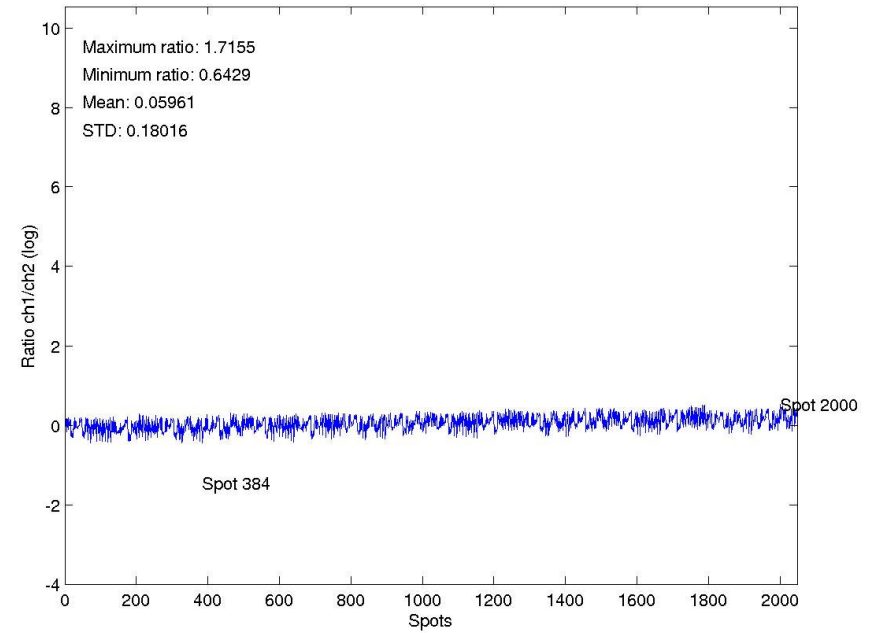




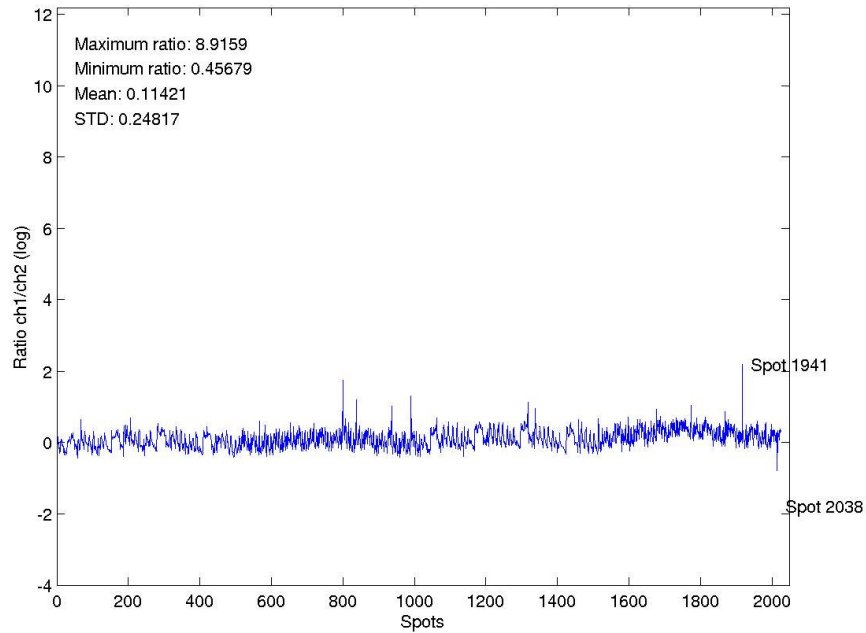
GHE037 2 CCD Quantarrayhist. Ratio ch1/ch2 (log)



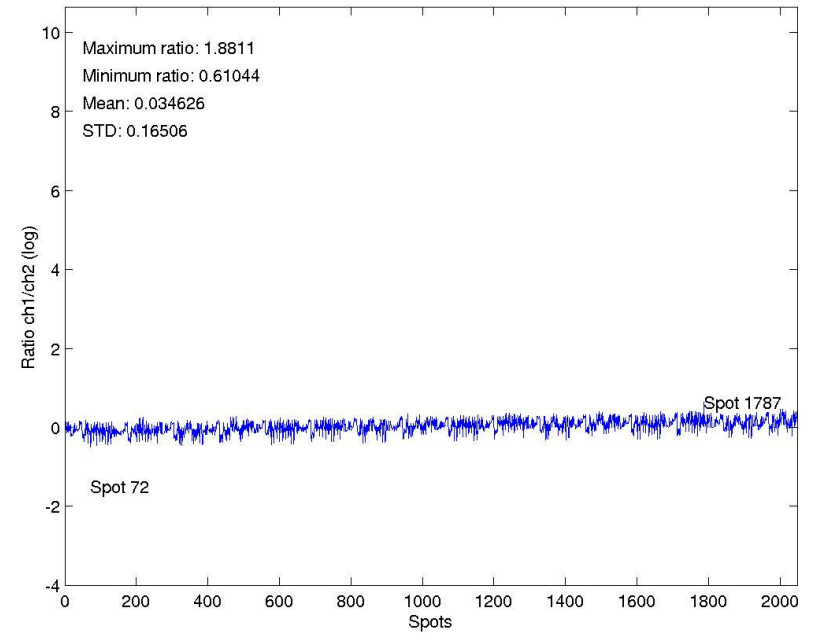
GHE038 Genepix 1st Quantarrayhist. Ratio ch1/ch2 (log)



GHE037 2 CCD IMEarray med. Ratio ch1/ch2 (log)

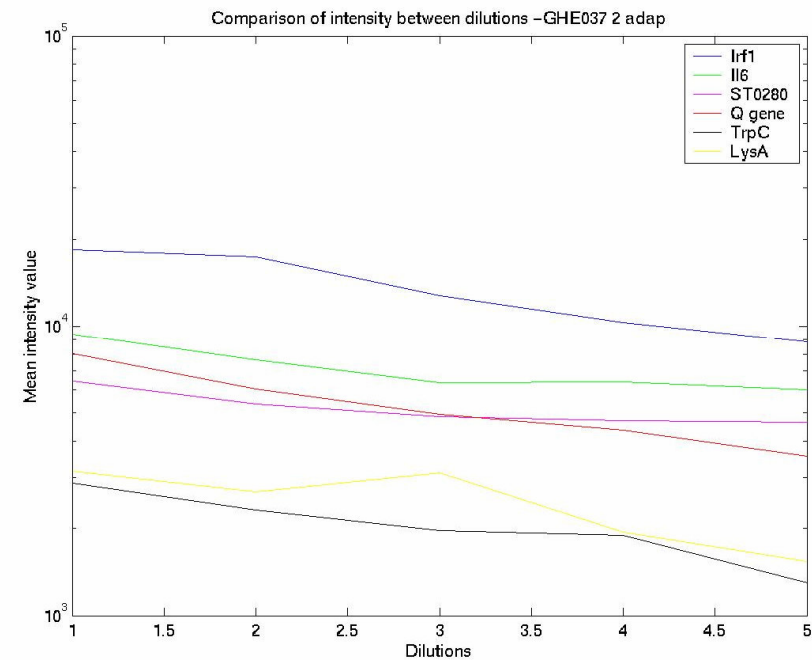
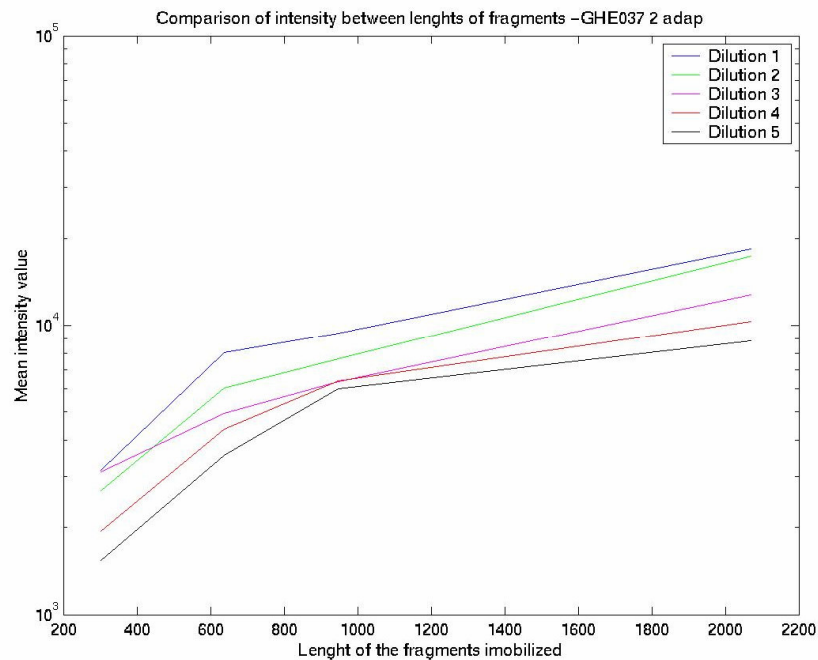


GHE038 Genepix IMEarray med. Ratio ch1/ch2 (log)



Parâmetros que influenciam o sinal

Tamanho do fragmento immobilizado: 700bp?



Hibridização com excesso de "Target"

O que nós fizemos:
Tamanho do frag. imob.



Target



O que eles fizeram:
Tamanho do frag. imob.



Target



Stillman et al, Analytical Biochemistry 259:149-157(2001)

Próxima etapa:
Tamanho do frag. imob.



Target



Próximos Passos...

Os desenhos dessas lâminas nos permitem avaliar muitas outras características dos experimentos de *arrays*

Exemplo:

Simular perfis de expressão entre esses 6 cDNAs, visando observar a capacidade de detecção de genes muito ou pouco expressos



Luiz F. L. Reis

Adriana Dias

Alex Carvalho

Beatriz Stolf

Chamberlein Neto

Eduardo Abrantes

Franco B. Runza

Lara Termini

Luciana Gomes

Regina Nomizo

Ricardo Alves

Sibele I. Meireles

Suzana Diniz

Waleska Martins

Junior Barrera

Daniel Dantas

Eduardo Jordão Neves

Roberto Hirata