BIOS FOR EVER

Carlos Eduardo Pedreira
COPPE PESC
Área de Inteligência Artificial

Yesterday

Apparatus and method for detecting cancer in tissue

US 3789832 A

RESUMO

An apparatus and method in which a tissue sample is positioned in a nuclear induction apparatus whereby selected nuclei are energized from their equilibrium states to higher energy states through nuclear magnetic resonance. By measuring the spin-lattice relaxation time and the spin-spin relaxation time as the energized nuclei return to their equilibrium states, and then comparing these relaxation times with their respective values for known normal and malignant tissue, an indication of the presence and degree of malignancy of cancerous tissue can be obtained.

Número da publicação	US3789832 A
Tipo de publicação	Concessão
Data de publicação	5 fev. 1974
Data de depósito	17 mar. 1972
Data da prioridade ⓘ	17 mar. 1972
Também publicado como	CA1004297A1
Inventores	Damadian R
Cessionário original	Damadian R
Exportar citação	BiBTeX, EndNote, RefMan

Citações de patente (3), Citações de não patente (1), Citada por (70), Classificações (9), Eventos legais (2)

Links externos: USPTO, Cessão do USPTO, Espacenet

In 1977 it takes place the first MRI in humans. It took 5 hours to generate the image.

The first commercial device is produced in 1980.

Here, There and Everywhere

Flow Citometry Data Analysis

Flow Cytometers are essential instruments for the **diagnosis** and follow up of a wide spectrum of diseases, mainly including **HIV-infection**, **leukemias** and **lymphomas** .

In the early 70's, the company Becton Dickinson put on the market the first flow cytometers

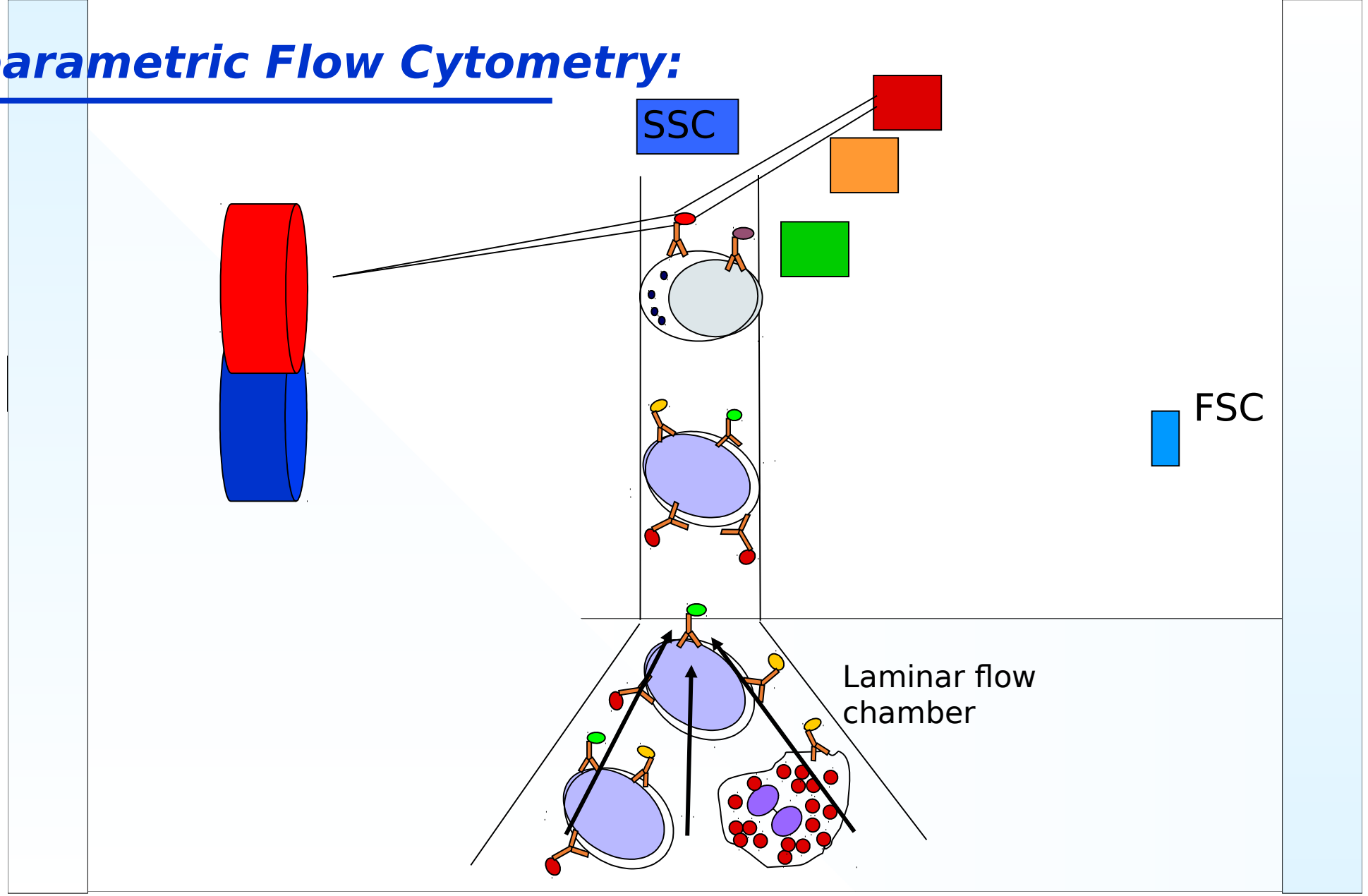
1 to 2 fluorescence detectors

Current Model

3 to 4 fluorescence detectors

8 fluorescence detectors

Multiparametric Flow Cytometry:



Flow Cytometers are able to perform fast evaluation of **multiple parameters in millions of cells.**

Accordingly, information is accessed **for each measured cell**.

A **HUGE amount of data** is being routinely generated, enhancing the need to **process** these data in a INTELLIGENT way to extract the desired information.

¡ Big Bio Data !

Help

ANOTHER PROBLEM: PROTEINS IDENTIFICATION

↓ *~40 000 proteins*

↔ *51 patients and 8 healthy controls* ↔

			Pacientes ---->	1	2	3	4	5	6	7
				PAT 7657	PAT 7938	PAT 7942	PAT 8014	PAT 8015	PAT 8062	PAT 8063
Evolución ---»»			Al diagnóstico ->	Metastásicos	Metastásicos	Metastásicos	Metastásicos	No Metastásicos	No Metastásicos	Metastásicos
			Evolución ->	1	1	1	1	2	2	1
			Final ->	1	1	1	1	2	2	1
proteinas		p-value ↓								
TEX11	1	0,0016286 3	1961	2,1555696	2,5947814	1,3210901	1,1990546	0,63505673	4,066673	0,4553764
BHMT2	2	0,0019947 8	1815	1,5596102	2,5012817	1,125496	1,1829665	0,42764947	3,091407	0,33466455
STC2	3	0,0019947 8	1945	1,6529819	3,43022	1,4345645	1,6283025	0,79565376	4,0478544	0,43871948
D21S2056 E	4	0,0023402 1	1816	1,5794747	2,4528308	1,1935892	1,0326964	0,4383045	2,813875	0,35407746
GTF2H1	5	0,0023402 1	1817	1,6336178	2,4554389	0,9736833	1,0083657	0,3445678	2,7225344	0,31672812
PSME3	6	0,0023402 1	1964	1,7977356	2,9674377	1,3902018	1,3800634	0,48554277	3,103187	0,3718307

...

Which proteins may differ 'healthy' from 'pathological' ?

Which proteins may differ 'metastatic' from ‘non metastatic’?

Which proteins may predict ‘evolution’?

↓ ~40 000 proteins

↔ 51 patients and 8 healthy controls ↔

			Pacientes ---->	1	2	3	4	5	6	7
				PAT 7657	PAT 7938	PAT 7942	PAT 8014	PAT 8015	PAT 8062	PAT 8063
Evolución	---»»		Al diagnóstico ->	Metastásicos	Metastásicos	Metastásicos	Metastásicos	No Metastásicos	No Metastásicos	Metastásicos
			Evolución ->	1	1	1	1	2	2	1
			Final ->	1	1	1	1	2	2	1
proteinas		p-value ↓								
TEX11	1	0,0016286	1961	2,1555696	2,5947814	1,3210901	1,1990546	0,63505673	4,066673	0,4553764
BHMT2	2	0,0019947	1815	1,5596102	2,5012817	1,125496	1,1829665	0,42764947	3,091407	0,33466455
S[illegible]C2	3	0,0019947	1945	1,6529819	3,43022	1,4345645	1,6283025	0,79565376	4,0478544	0,43871948
D21S2056E	4	0,0023402	1816	1,5794747	2,4528308	1,1935892	1,0326964	0,4383045	2,813875	0,35407746
GTF2H1	5	0,0023402	1817	1,6366174	2,4554389	0,97566533	1,0[illegible]857	0,34456784	2,[illegible]554	0,31672812
PSME3	6	0,0023[illegible]	1964	1,7977356	2,9674377	1,3902018	1,3800634	0,48554277	3,103187	0,3718307

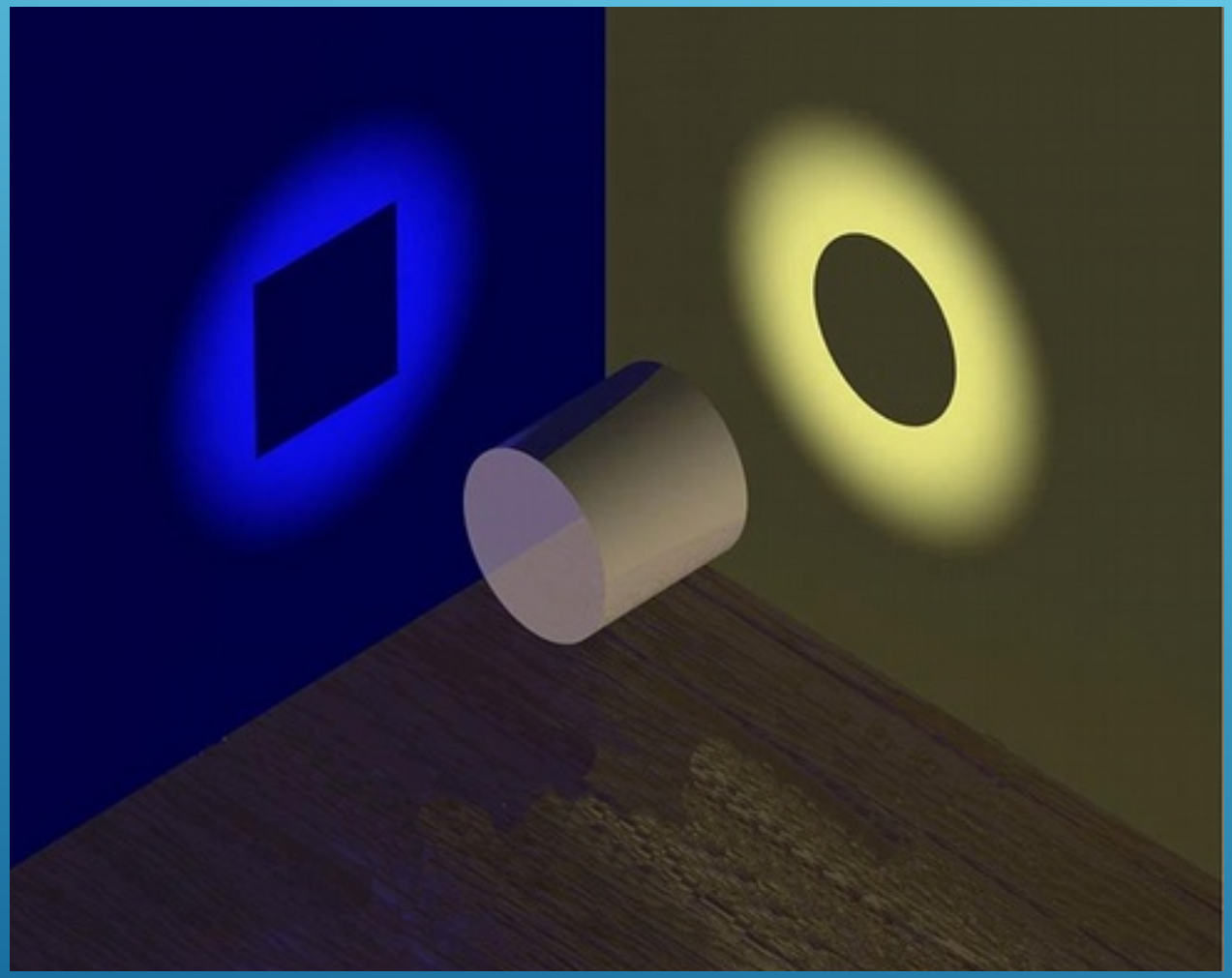
...

Note that we have THOUSANDS of attibutes and few observations

Here Comes the Sun

Projecting in 2-D

The way one projects = The way one sees



Why (and when) one should project in 2D aiming classification?

Why:

- Frequently, one needs a decision support tool and not an automatic classification algorithm. **The final decision is to be taken by the user, not by the 'system'**.

When:

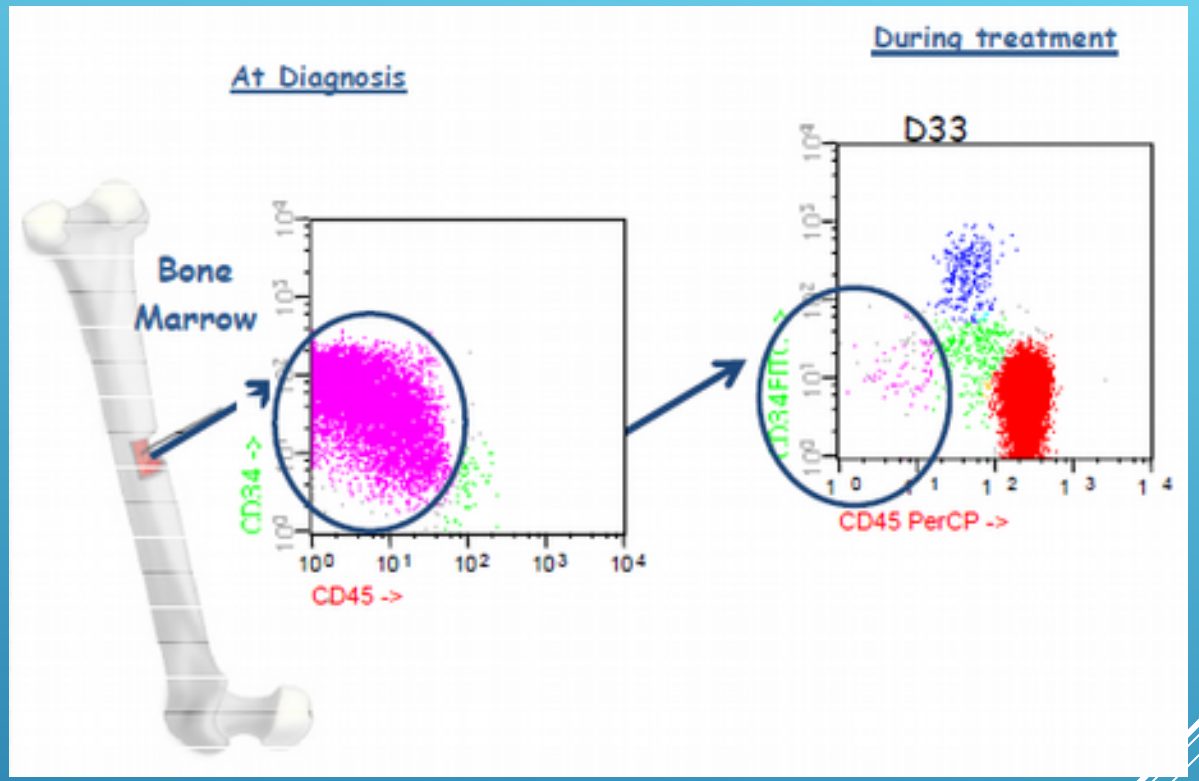
- One does not want to classify in automatic mode by **ethical or legal reasons** e.g. medical diagnosis.
- One has **additional individualized information** that is difficult to model but relevant to be added.

Get Back

BACK TO CYTOMETRY DATA

Minimal Residual Disease (MRD)

12 attributes per cell, of 5 million cells



- MRD is a **prognostic factor** in several hematological diseases.
- MRD is a **criteria to change treatment strategies** in several hematological diseases.

diagnostic

Almost all cells are pathological

treatment

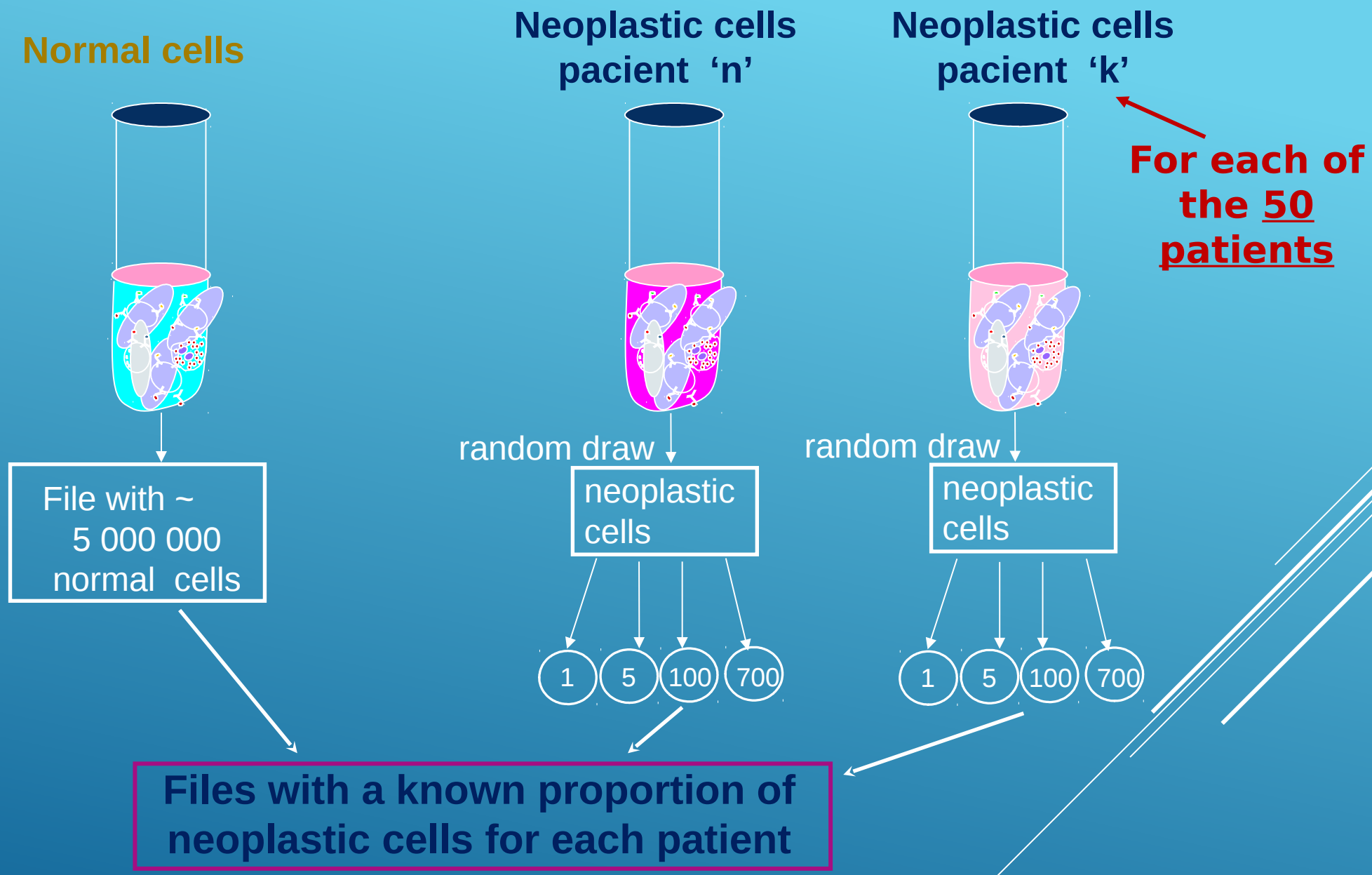
Mainly normal but residual pathological cells may be present

Patholocical cells?

yes

How many ?

Testing the sensitivity of the method



Consequently, **for each of the 50 patients**, **88 “MRD-files” were generated containing known proportions of** between 1 and 1000 **neoplastic B cells** in the pool of 5 x 10^6 normal cells

Every Little Thing

Results

Sensitivity:

In 80 % of the cases (**40/50**), the method was able to detect just 1 patological event in **5×10^6** normal cells.

Level of agremment:

For 90% of the pacients (**45/50**), the **correlation coeficient (r^2) was greater than 0.999**. The other 10% (5cases) reached **$0.964 \leq r^2 \leq 0.999$**.

Differential diagnosis

Goal:

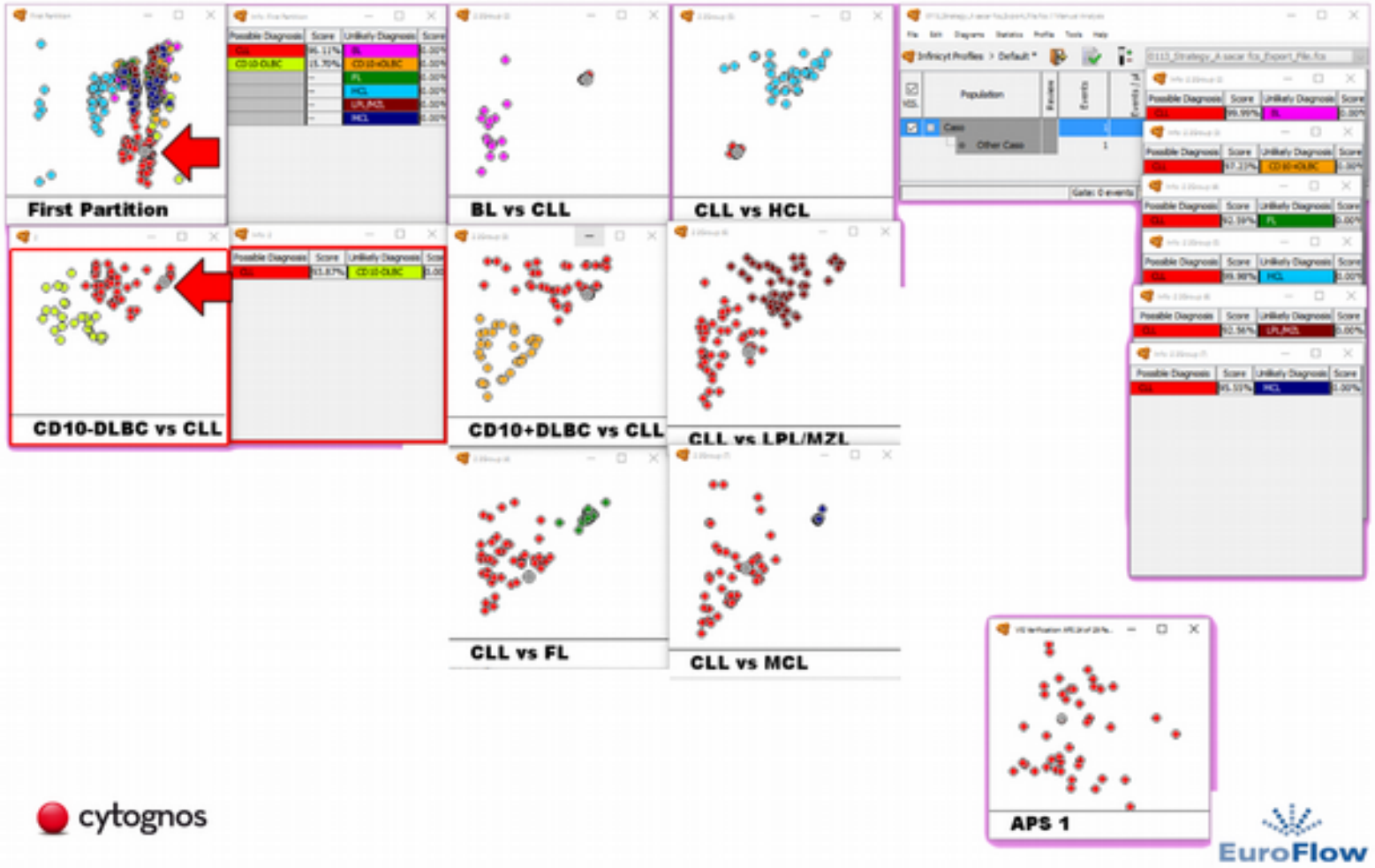
To differentiate, using flow cytometry data, **among 8 types of lymphomas**: BL, CD10-, CD10+, CLL, FL, HCL, MCL, LPL-MZL

Here, we use the **mean in the 24 attributes** for each patient. The goal is to **differentiate among patients** and not among cells of a single patient.

2-D projection, the final decison is taken by the user

- The cost function aims to preserve the **distance** structure of the **observations to pre-stablished prototypes** (representing the classes).
- Furthermore, we model the probability (in R^2) of any observation (patient) to belong to any of the classes (type of Lymphoma).
- The attributes are re-selected at each step (so that the spaces change).
- Probability thresholds are created to provide a hierarchical scheme.

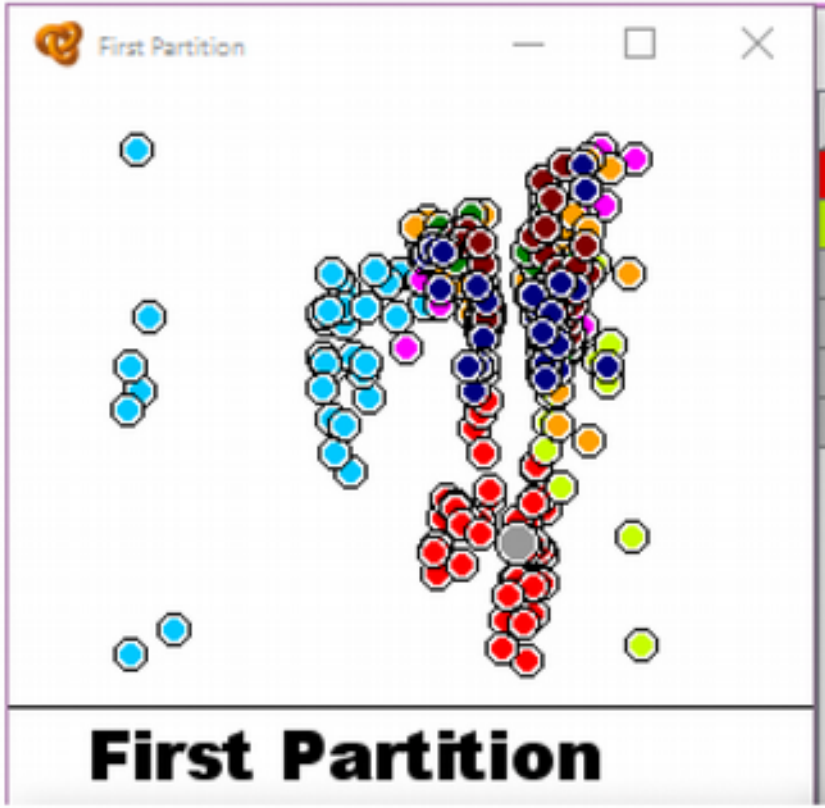
Modelagem Hierárquica (no software Infinicyt)



cytognos

EuroFlow

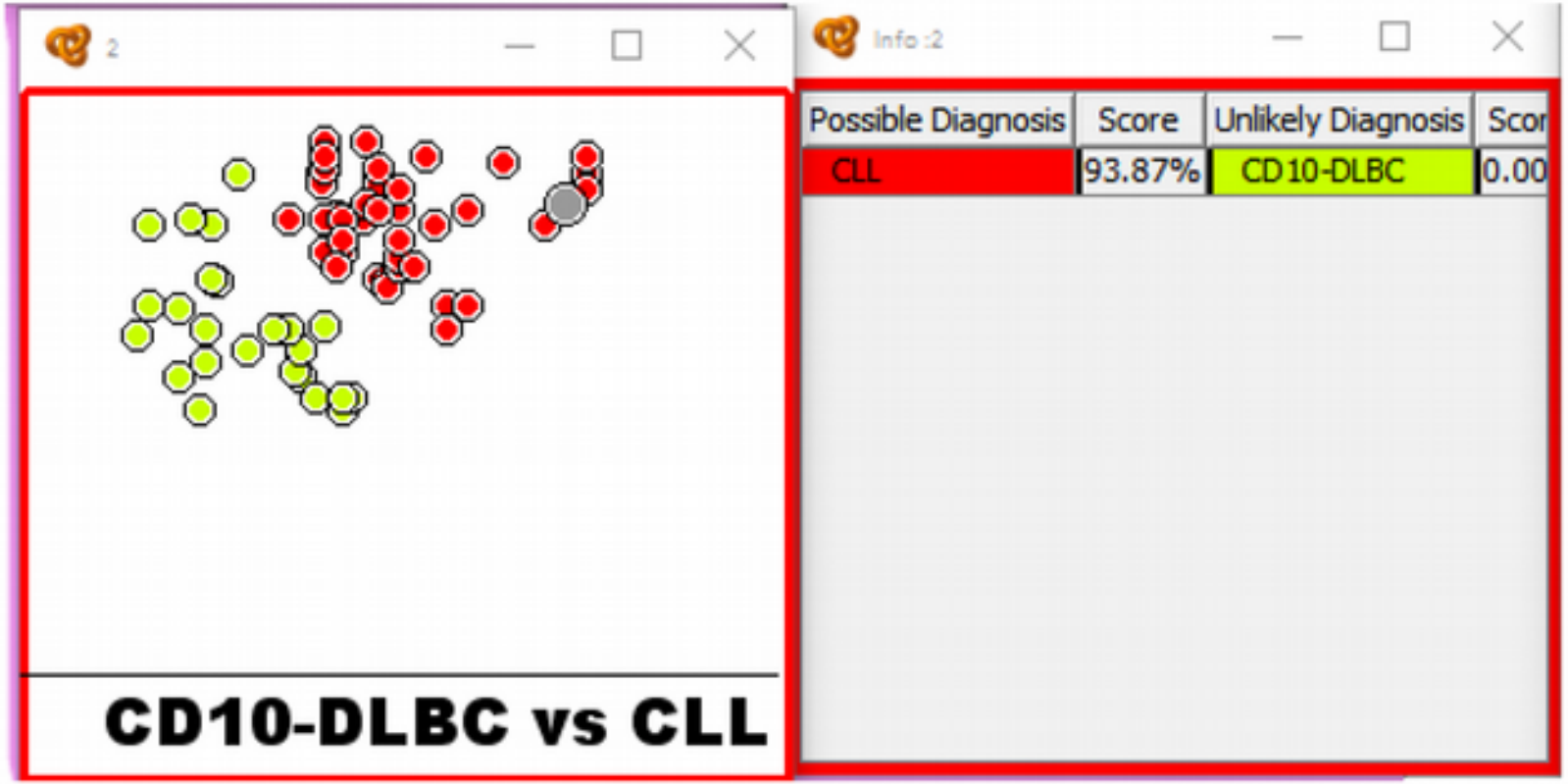
Modelagem Hierárquica (no software Infinicyt)



Info :First Partition

Possible Diagnosis	Score	Unlikely Diagnosis	Score
CLL	96.11%	BL	0.00%
CD10-DLBC	15.70%	CD10+DLBC	0.00%
	--	FL	0.00%
	--	HCL	0.00%
	--	LPL/MZL	0.00%
	--	MCL	0.00%

Modelagem Hierárquica (no software Infinicyt)



The Long and Winding Road

From Academic research to Real world application

Academic:

- **Pedreira, C.E.**; Costa,E.S; Lecrevisse Q.; van Dongen J.J.M.; Orfao A. "Overview of Clinical Flow Cytometry Data Analysis: Recent Advances and Future Challenges" **Trends in Biotechnology**, Vol 31 n.7, pp415-427, (2013).
- Costa ES; **Pedreira CE**; Flores J; Lecrevisse Q; Quijano S; Barrena S; Almeida, J; Böttcher S; Van Dongen JJM; Orfao A; "*Automated Pattern-Guided Principal Component Analysis versus Expert-Based Immunophenotypic Classification of Hematological Malignancies*" **Leukemia** , 24(11):1927-33, (2010).
- **Pedreira CE;** Costa ES; Arroyo ME; Almeida J; Orfao A; "A Multidimensional Classification Approach for the Automated Analysis of Flow Cytometry Data"; **IEEE Transactions on Biomedical Engineering**, Vol 55, p.1155-1162; (2008).

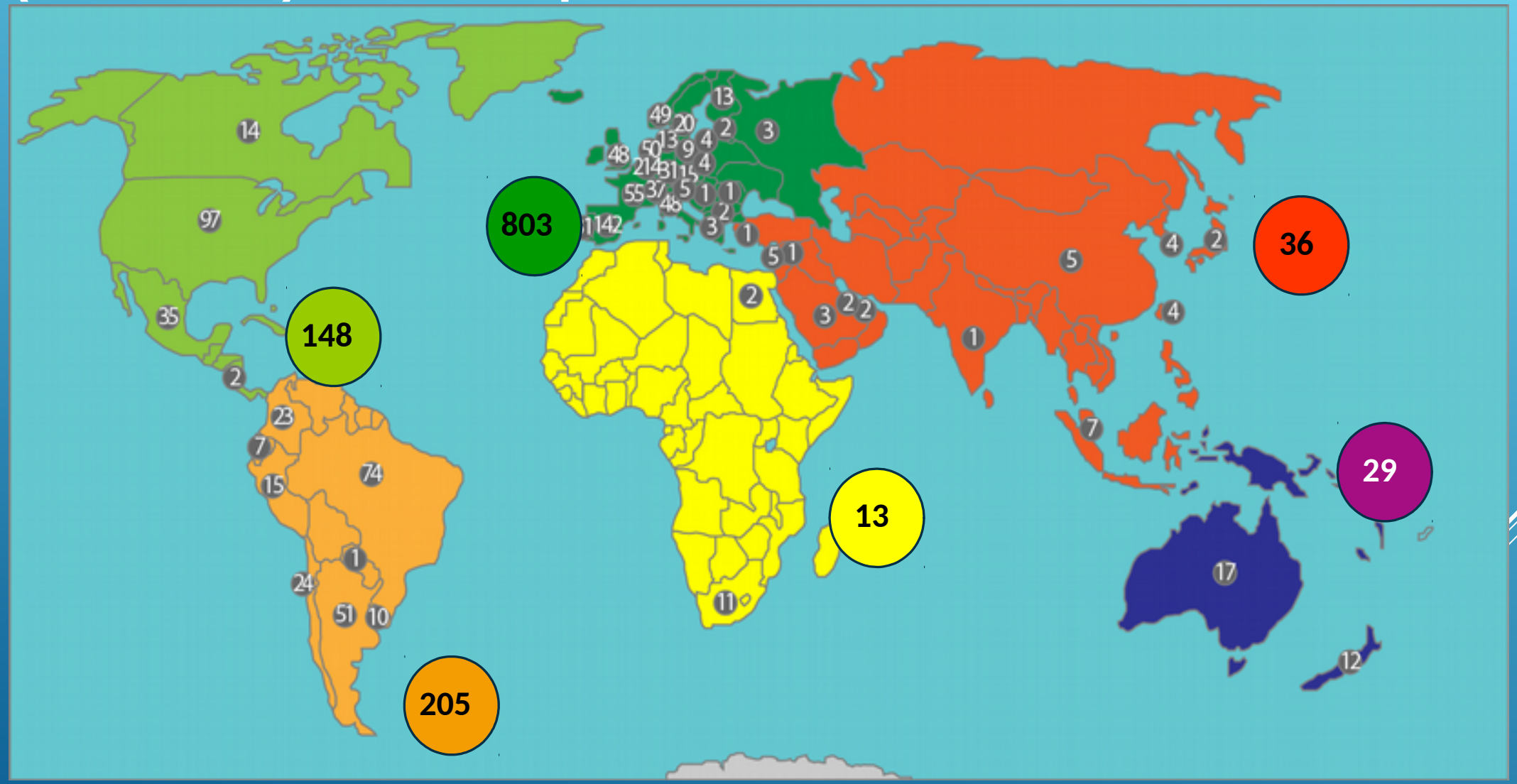
Inovation:

- **United States Patent nº US 7,321,843B2** "Method for generating flow cytometry data files containing an infinite number of dimensions based on data estimation" (concession 2008). Inventors: Alberto Orfao de Matos, **Carlos Eduardo Pedreira** and Elaine Sobral da Costa. **License assigned to Becton/Dickinson Biosciences and Cytognos SL**.
- **Internacional Patent nº WO 2010/140885 A1** (Provisional) "Methods, reagents and kits for flow cytometric immunophenotyping" (December 2010). Inventors: JJM van Dongen, A Orfao, JA Flores, JM Parra, VHJ van der Velden, S Bottcher, AC Rawstron, RM de Tute, LBS Lhermitte, V Asnafi, E Mejstrikova, T Szczepanski, PJ Lucio, M Ayuso, **CE Pedreira**. **License assigned to Becton/Dickinson Biosciences and to Cytognos SL**.

IN USE (making knowlegde avaliable in the real world):

- **Software** *' INFINICYT'* uses our results (patents and papers). It is a key tool for cytometry, including leukemia and lymphomas diagnosis and follow up. It is currently **licensed and in day-to-day use in more than 50 countries**. It is considered to be the state-of-the-art software for analysis and interpretation of flow cytometry data.

EuroFlow / Infinicyt users (2009-2016): ~1234 copies sold in all continents



September 2016

All We´ve Got To Do

Future Perspective Computational Modeling in Medicine

- **Data mining tools will gain more and more play a key role in** extracting relevant information in an objective, precise, reproducible and comprehensive way.
- Information should be **made available to users** through **intuitive graphical representations** and user-friendly interpretation-guided tools.
- The **avalanche of medical data** will continue to push for **quantitative tools**.

*Some of the **frontier problems in healthcare** will be tacked by a **new generation of professionals** capable of absorbing different technologies and who will be able to work side by side with colleagues with **distinct backgrounds** in engineering, statistics computing and health sciences.*

Come Together

Close partners

*Some of these ideas and results are part of the my investigation within the **EuroFlow consortium**. UFRJ is the only non-European group in this consortium and the main responsible for the **data analysis** developments.*

*Part of the developments are done in association we the **UFRJ Pediatric Hospital Cytometry Lab in Rio (IPPMG)** where we maintain a lab.*

EuroFlow
a division of ESLHO

The EuroFlow Group

EuroFlow
a division of ESLHO

With a Little Help From my Friends

My main colaborators (Students and non COPPE):

- Profa. Elaine S. Costa (Faculdade de Medicina- UFRJ)
- Prof. Alberto Orfao Univ. de Salamanca, Espanha
- Prof. Manoel Fuentes Univ. de Salamanca, Espanha
- Prof. Rodrigo Peres CEFET
- Diego, Laura, Lygia, Luciana,

- And of course: John, Paul, George & Ringo

We Can Work It Out

Thank You !

www.cos.ufrj.br/~pedreira

pedreira56@gmail.com