Reutilización de datos para investigación biomédica

Potentiating the role of “Data Scientist”

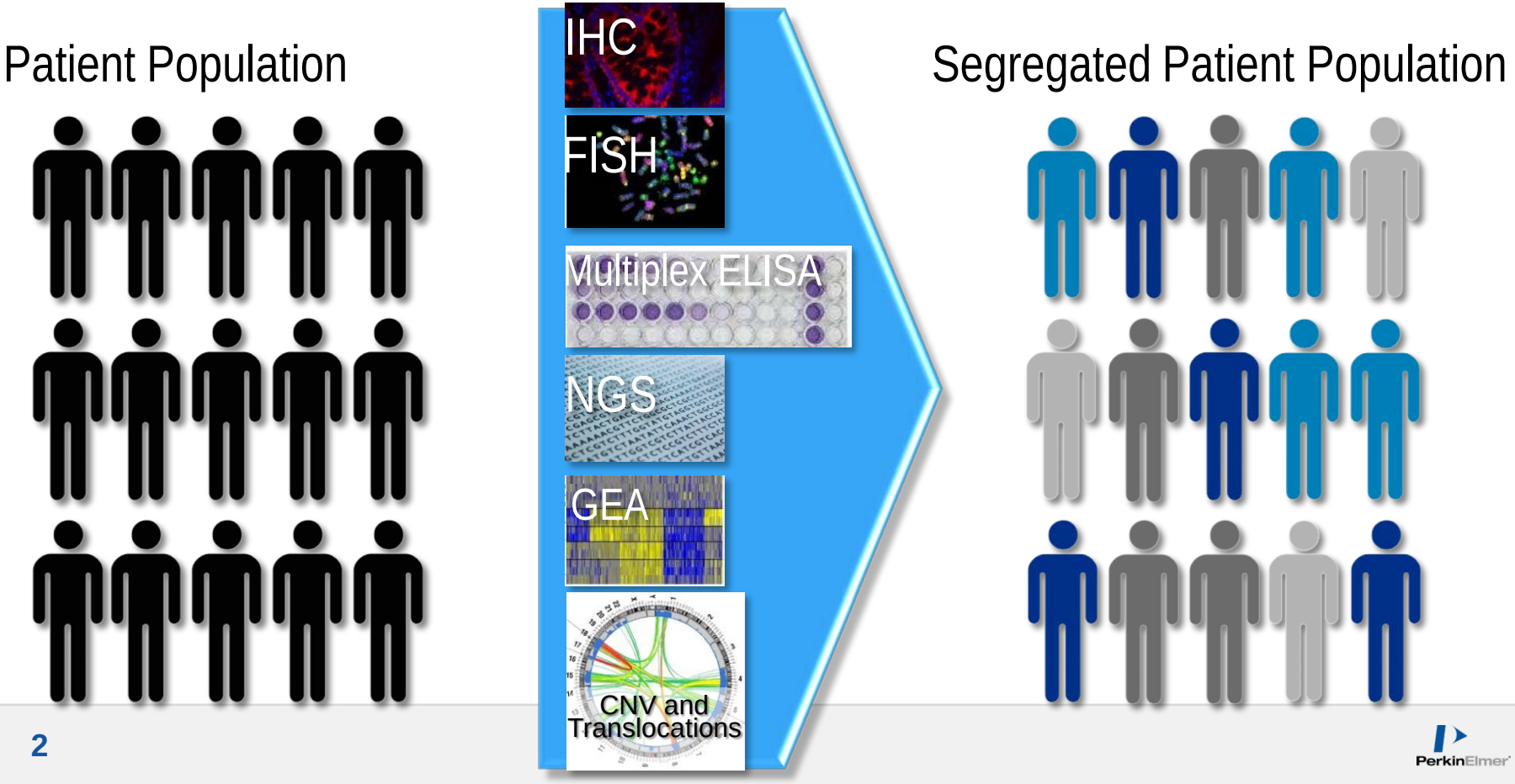
Eduardo Gonzalez-Couto, PhD

March, 24th 2015

PerkinElmer
For the Better

The biomedical industry vision

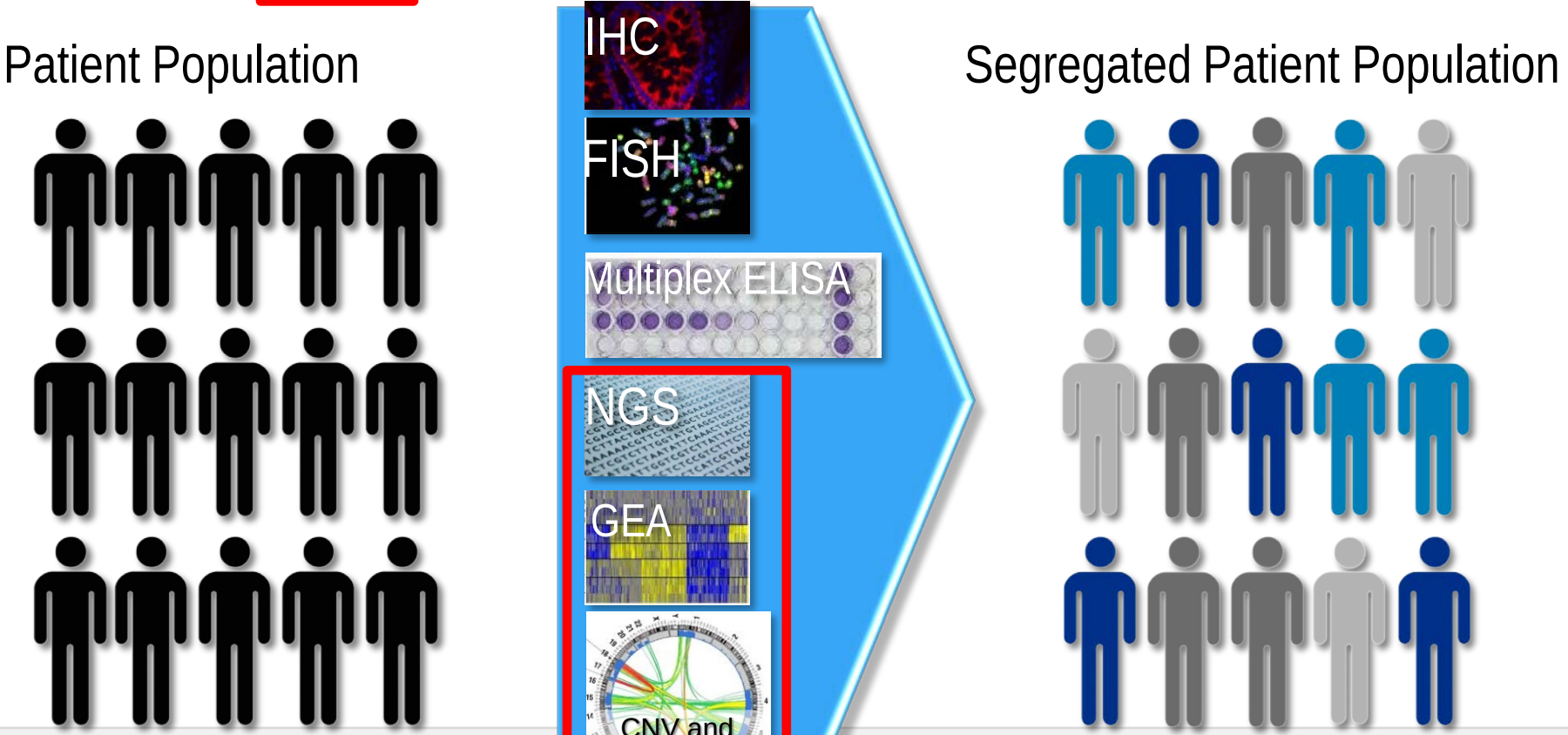
- Translational Medicine is already transforming how new therapies and devices are discovered and developed



PerkinElmer

The biomedical industry vision

- Translational Medicine is already transforming how new therapies and devices are discovered and developed and high-content Omics technologies are accelerating this trend



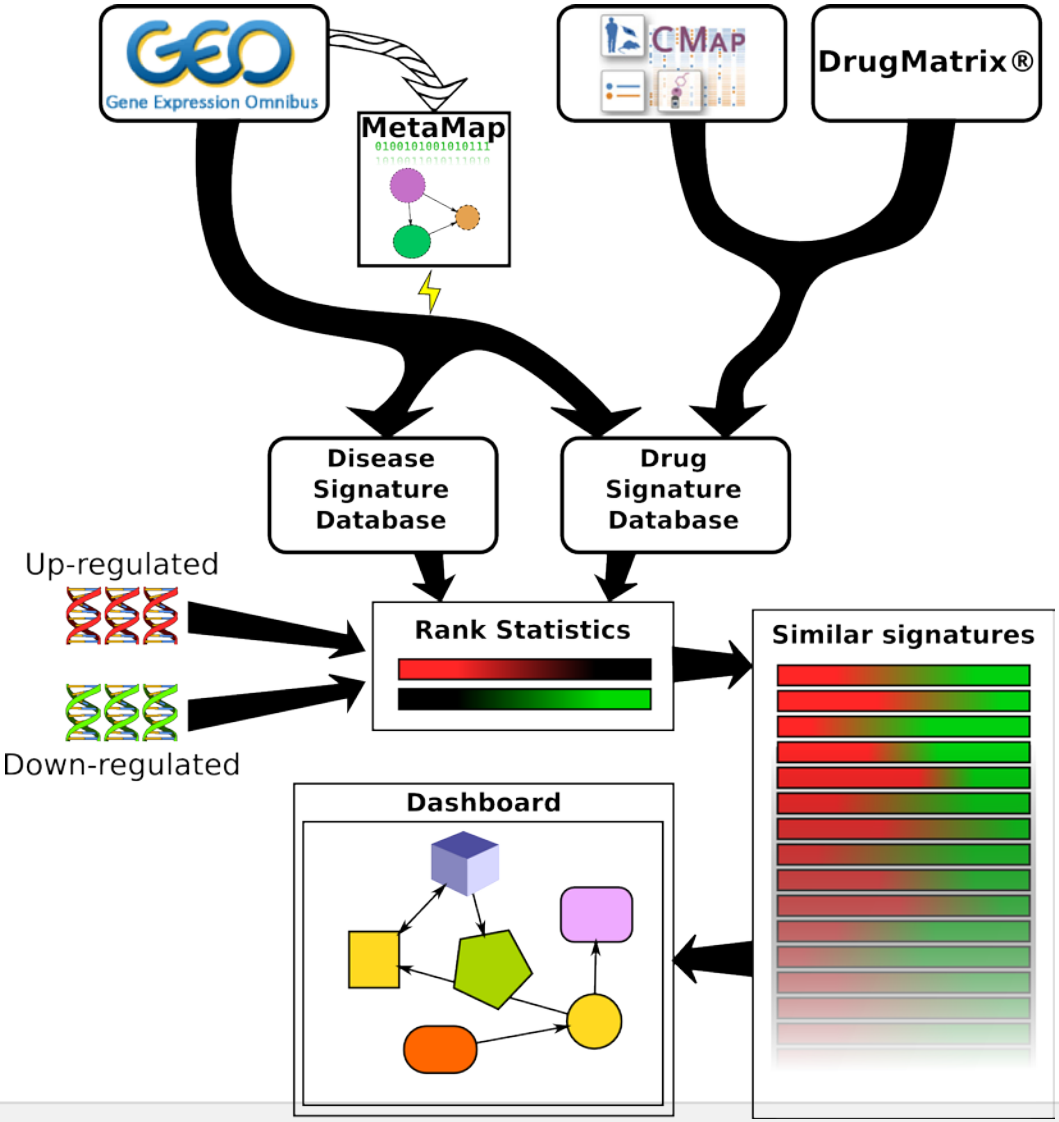
PerkinElmer

Data re-use in the biomedical industry

- Data re-use examples in the biomedical industry
  - Gene Expression Omnibus (GEO) – Children's Tumor Foundation
  - SciDB - Novartis
  - The Cancer Genome Atlas (TCGA) - Roche
- Devices re-use new trend
- Potentiating a new role

Data re-use in the biomedical industry

- Gene Expression Omnibus

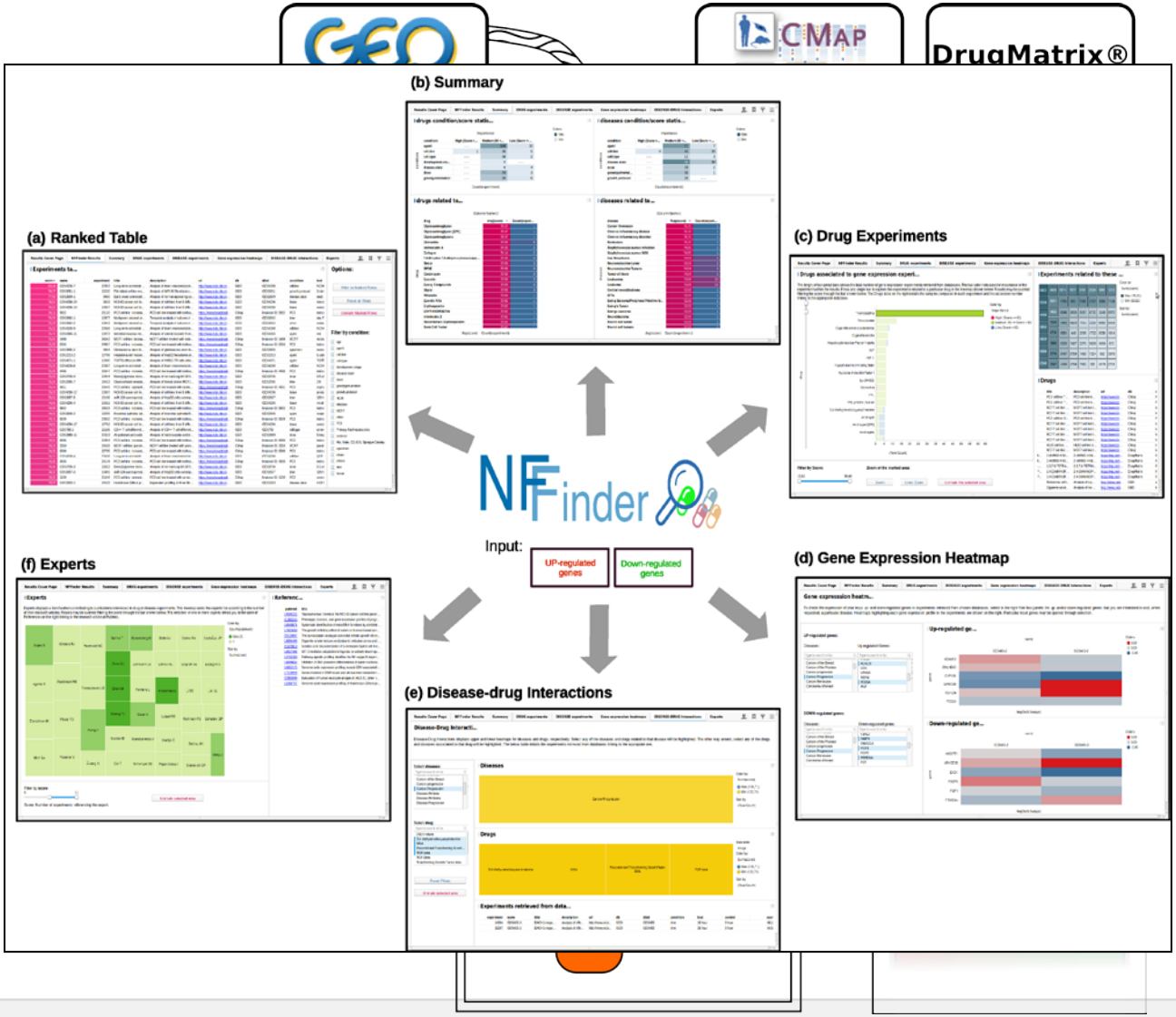


CHILDREN'S TUMOR FOUNDATION
ENDING NF THROUGH RESEARCH

PerkinElmer

Data re-use in the biomedical industry

- Gene Expression Omnibus



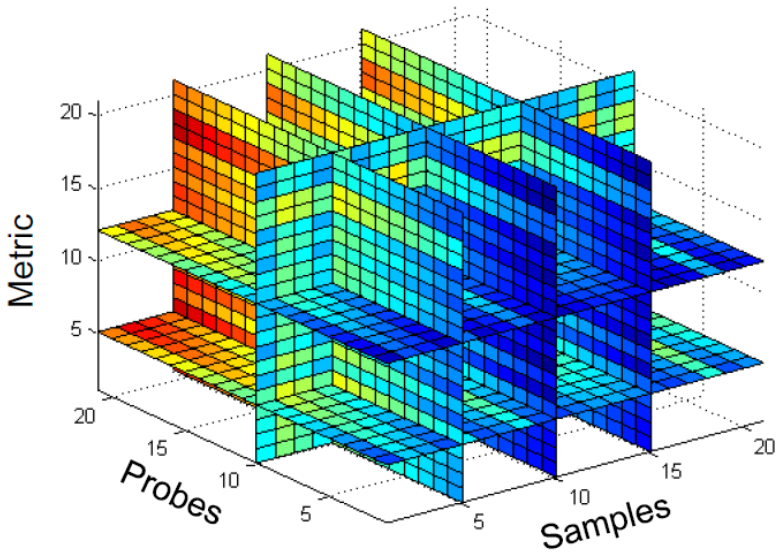
CHILDREN'S TUMOR FOUNDATION
ENDING NF THROUGH RESEARCH

PerkinElmer

Data re-use in the biomedical industry

- SciDB - Novartis
  - Microarray – platform for simultaneously measuring the expression levels of thousands of genes
  - 60,000+ arrays
  - 60,000 probes (multiple probes per gene)
  - >100 indications
  - 6B+ data points
  - Use cases:
    - Target finding
    - Patient stratification
    - Biomarker identification

SciDB
Developed by Paradigm4

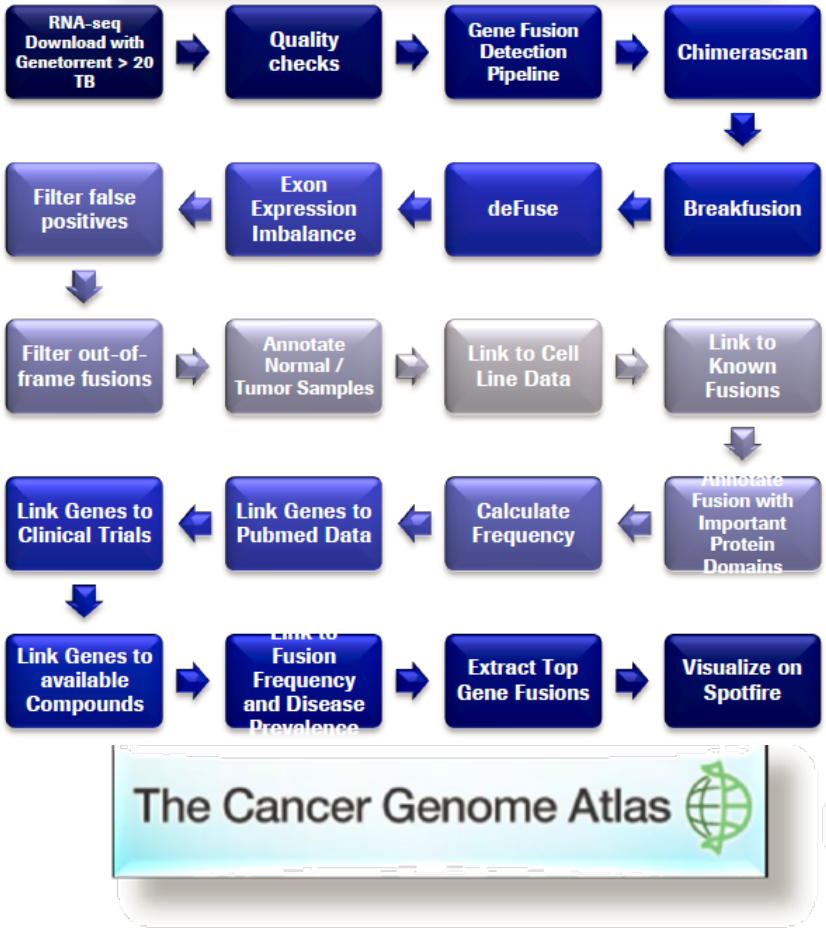


NOVARTIS

PerkinElmer

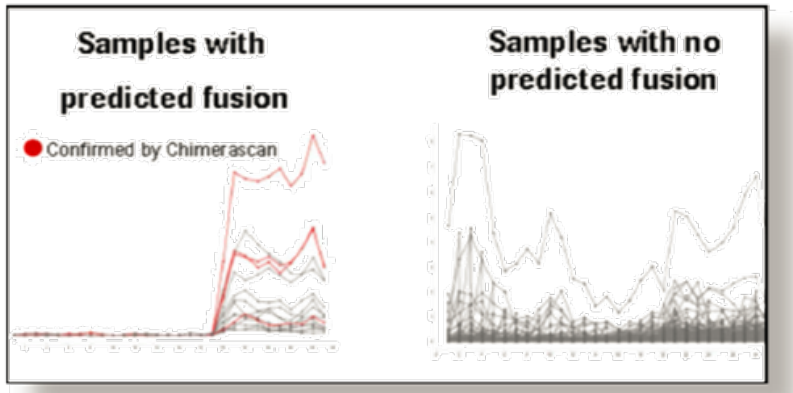
Data re-use in the biomedical industry

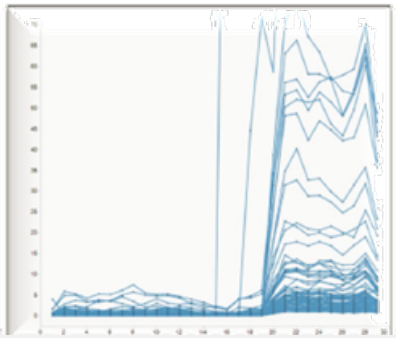
- The Cancer Genome Atlas



Roche

Integration with exon expression imbalance algorithm reveals additional gene fusions





Similar patterns in other disease

- **Confirmed by IHC**
- **Not previously described**
- **Possibly new disease indication**

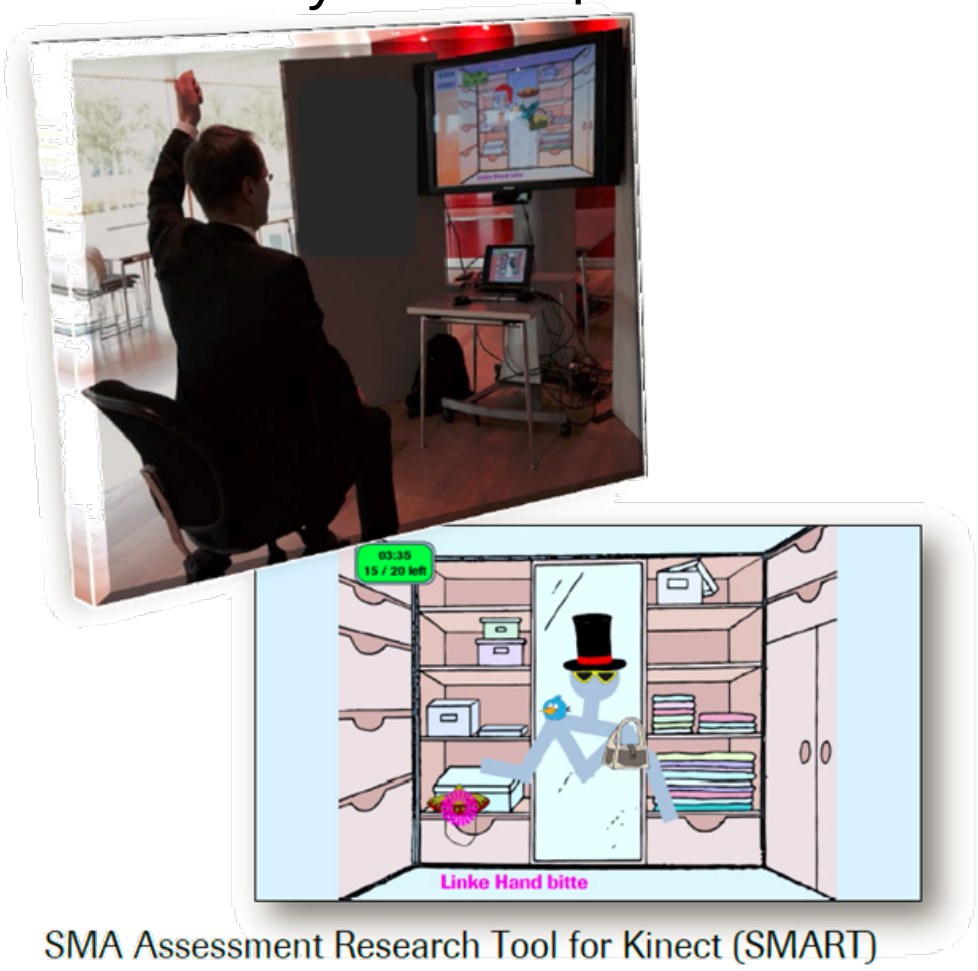
PerkinElmer

Re-use in the biomedical industry

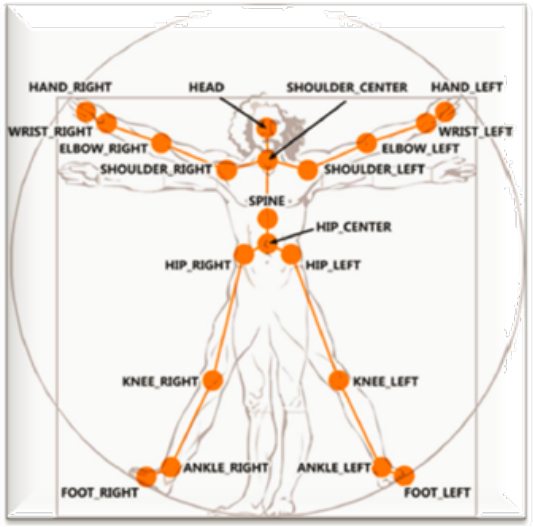
- Not only re-using the data…

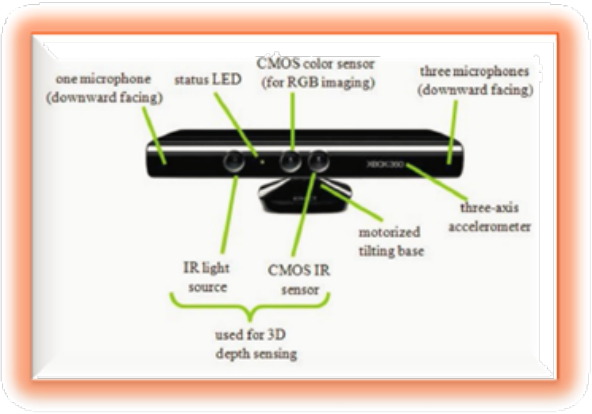
Devices re-use new trend

- Sensor analysis for improved trial designs



SMA Assessment Research Tool for Kinect (SMART)





PerkinElmer

Potentiating a new role

- Challenge
  - The complexity and size of the data, coupled to complex technologies limit the opportunity for life science experts to explore and interpret the data

- A solution – Data Science
  - Integrate the tools allowing to process the data and visualise the results
  - Data Science combines strong scientific and disease domain expertise with analytics capabilities **to generate answers rather than information**

- Educate “Data Scientists” to be able to use such integrated tools, enabling them to perform advanced results exploration and queries

  - For instance finding patients with similar patterns of mutations in large genome-wide association studies databases

Gracias!

PerkinElmer