Data Visualization

Aik Choon Tan, Ph.D.

Associate Professor of Bioinformatics

Division of Medical Oncology

Department of Medicine

aikchoon.tan@ucdenver.edu

10/19/2018

http://tanlab.ucdenver.edu/labHomePage/teaching/BSBT6111/

Data Visualization

- Important communication skills to present data and provide clear "story"

- Help to find important patterns from vast amount of data

- Deliver key messages from "big data"

Data Visualization went Wrong
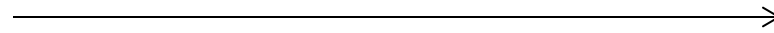
http://viz.wtf/

The Beauty of Data Visualization

http://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization?language=en

Visual Representation
Integral to Research Trajectory

Exploration   →   **Explanation**

Bench   Publication

Charts, graphs   *Networks*

Browsers   *Workflow*

Heatmaps   *Infographics*

Interactive Browsers

(Adapted from Bang Wong's slides)

Data Visualization in Practice

Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study

Pieter Mestdagh[1], Nicole Hartmann[2], Lukas Baeriswyl[2], Ditte Andreasen[3], Nathalie Bernard[4], Caifu Chen[4], David Cheo[5], Petula D'Andrade[6], Mike DeMayo[7], Lucas Dennis[8], Stefaan Derveaux[9], Yun Feng[5], Stephanie Fulmer-Smentek[6], Bernhard Gerstmayer[10], Julia Gouffon[7], Chris Grimley[8], Eric Lader[11], Kathy Y Lee[4], Shujun Luo[12], Peter Mouritzen[3], Aishwarya Narayanan[13], Sunali Patel[4], Sabine Peiffer[10], Silvia Rüberg[10], Gary Schroth[12], Dave Schuster[5], Jonathan M Shaffer[11], Elliot J Shelton[4], Scott Silveria[9], Umberto Ulmanella[4], Vamsi Veeramachaneni[13], Frank Staedtler[2], Thomas Peters[2], Toumy Guettouche[14] & Jo Vandesompele[1]
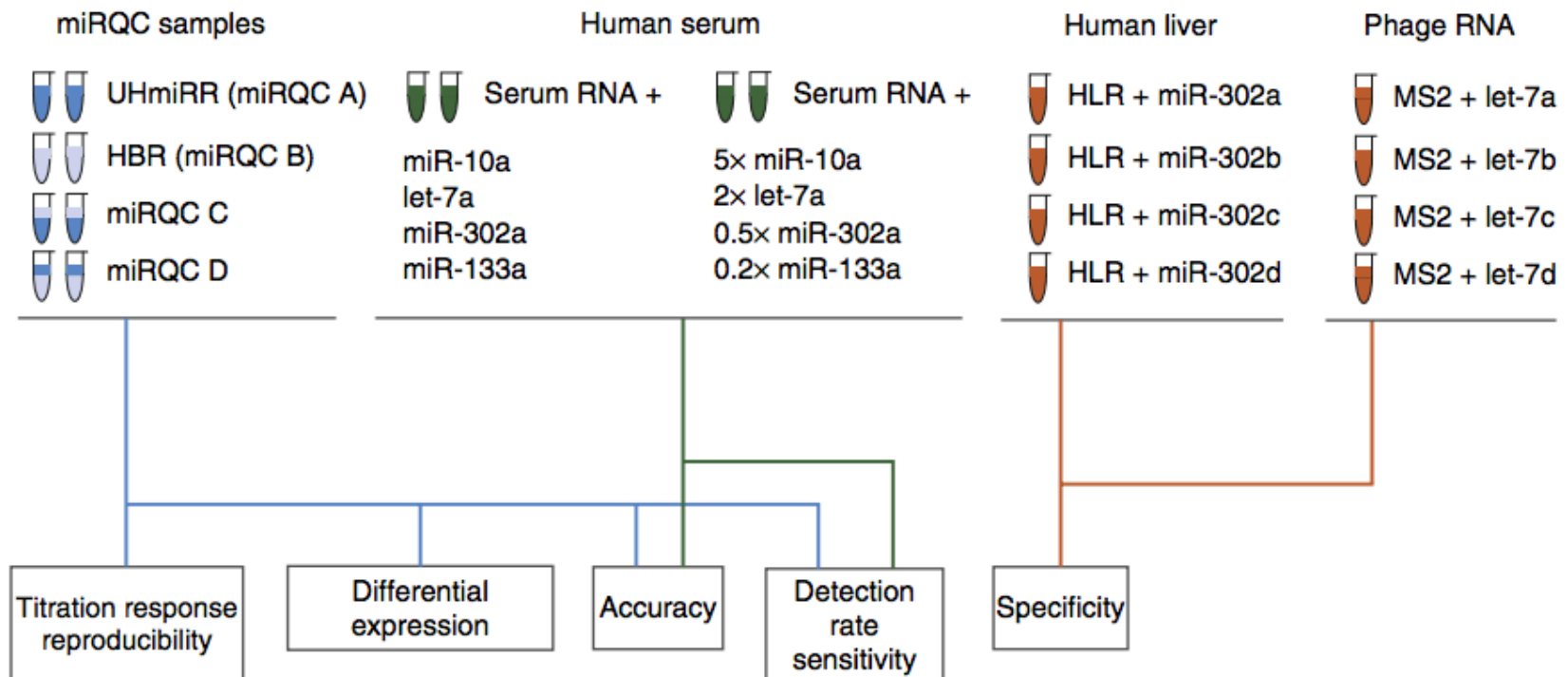
Nature Methods 2014
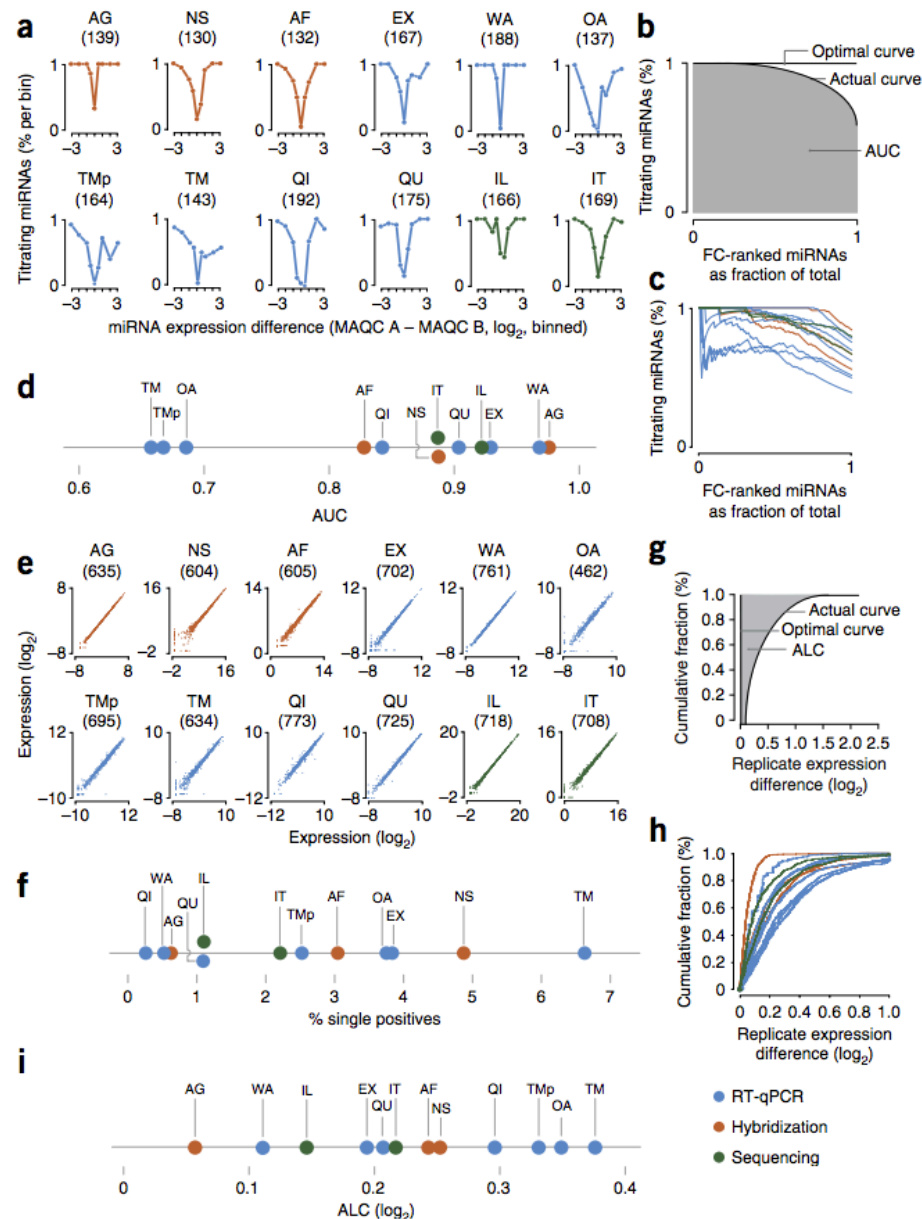
Data Visualization in Practice

Quantitative PCR (PCR)

EX   miRCury (Exiqon)
OA   OpenArray (Life Technologies)
TM   TaqMan Cards (Life Technologies) *
TMp TaqMan Cards preAmp (Life Technologies)
QI   miScript (Qiagen)
QU   qScript (Quanta BioSciences)
WA   SmartChip (WaferGen)

Hybridization (HYB)

AF   microarray (Affymetrix) *
AG   microarray (Agilent)
NS   nCounter (Nanostring) *

Sequencing (SEQ)

IL TruSeq (Illumina)
IT Ion Torrent (Life Technologies)

RNA input (ng)

EX	40	AF	400
OA	100	AG	100
TM	350	NS	100
TMp	50		
QI	500	IL	1,000
QU	800	IT	1,000
WA	1,000		

miRQC samples

UHmiRR (miRQC A)
HBR (miRQC B)
miRQC C
miRQC D

Human serum

Serum RNA +

miR-10a
let-7a
miR-302a
miR-133a

Serum RNA +

5× miR-10a
2× let-7a
0.5× miR-302a
0.2× miR-133a

Human liver

HLR + miR-302a
HLR + miR-302b
HLR + miR-302c
HLR + miR-302d

Phage RNA

MS2 + let-7a
MS2 + let-7b
MS2 + let-7c
MS2 + let-7d

Titration response reproducibility

Differential expression
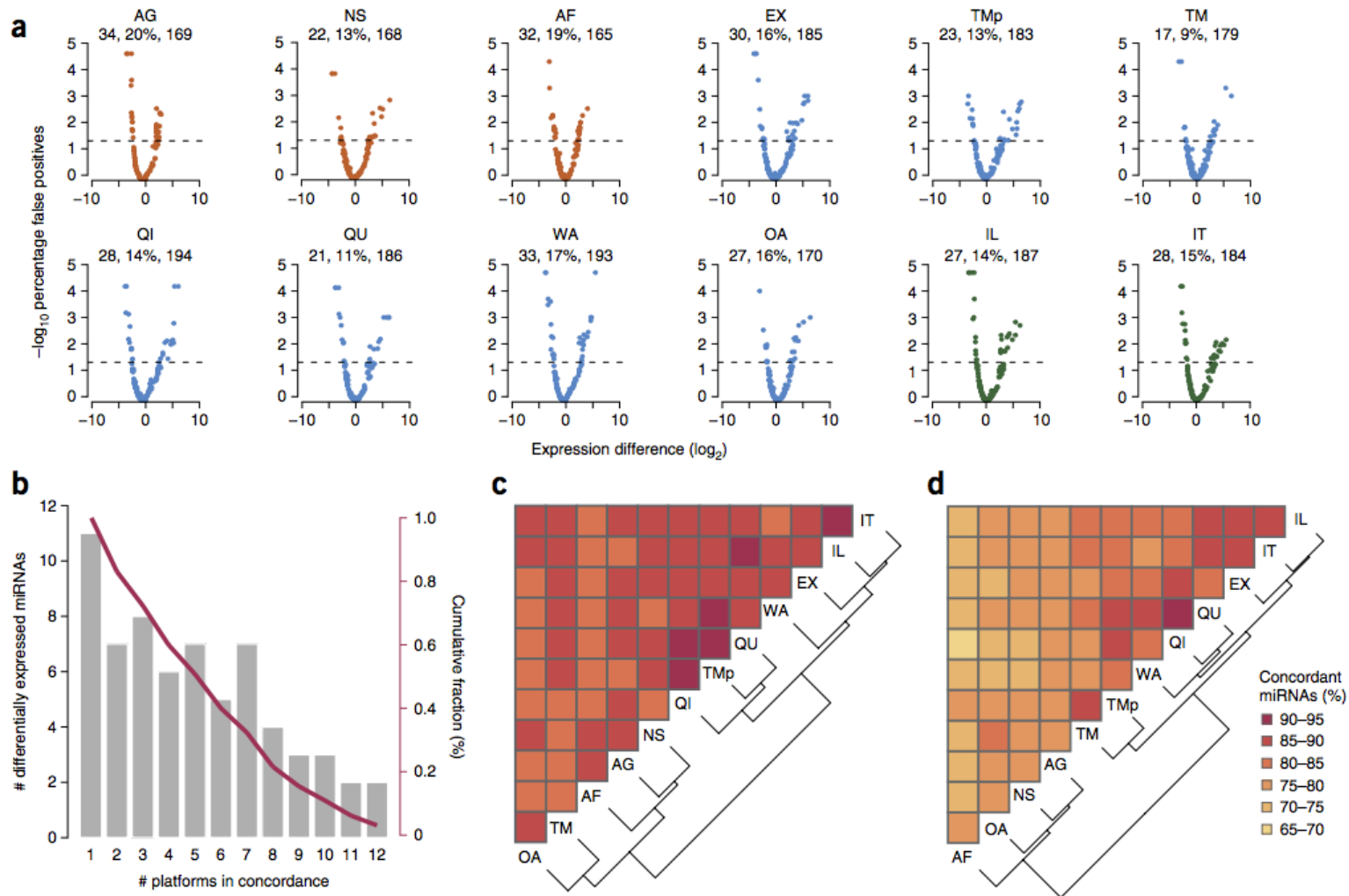
Accuracy

Detection rate sensitivity

Specificity

Data Visualization in Practice

Data Visualization in Practice

Figure 5 | Differential miRNA expression. (**a**) Volcano plot showing −log₁₀ of the rank products percentage of false positives value as a function of the mean expression difference for miRQC A and C (*n* = 4) versus miRQC B and D (*n* = 4) samples. For each platform (abbreviated as in **Fig. 1**), total number of differentially expressed miRNAs, is indicated above the plot as an absolute number and as a percentage relative to the total number of miRNAs included in the analysis, also indicated above the plot. (**b**) Number of differentially expressed miRNAs identified by at least one or multiple platforms. (**c**) Hierarchically clustered heatmap indicating miRNA concordance between all platform combinations. (**d**) Hierarchically clustered heatmap indicating miRNA concordance between all platform combinations, taking into account detection rate.

Charts & Graphs
(& Color & Design Principles)

- https://vimeo.com/114252896

- Bang Wong's Blog (Broad Institute)
  - Nature Methods – Points of View
  - http://clearscience.info/wp/?p=546

How to visualize Human Genome?

- ACTGACTG ..... (3 billion characters)
- Chromosome Map (NCBI Map Viewer)
- https://www.ncbi.nlm.nih.gov/projects/map view/map_search.cgi?taxid=9606&build=1 07.0

Circos Plot

- CIRCULAR VISUALIZATION
- Circos is a software package for visualizing data and information. It visualizes data in a circular layout — this makes Circos ideal for exploring relationships between objects or positions. There are other reasons why a circular layout is advantageous, not the least being the fact that it is attractive.
- http://circos.ca/

Circos Plot for Human Genome

http://circos.ca/intro/genomic_data/

Browser

- Integrative Genomics Viewer (IGV)
- The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.
- http://software.broadinstitute.org/software/igv/

cBioPortal

- The cBioPortal for Cancer Genomics provides visualization, analysis and download of large-scale cancer genomics data sets.

- http://www.cbioportal.org/

ICGC Data Portal

- The ICGC Data Portal provides tools for visualizing, querying and downloading the data released quarterly by the consortium's member projects.

- https://dcc.icgc.org/

NCI Genomic Data Commons

- The NCI's Genomic Data Commons (GDC) provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine.

- https://gdc-portal.nci.nih.gov/

COSMIC – Sanger Institute

- COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

- http://cancer.sanger.ac.uk/cosmic

GDSC – Sanger Institute

- GDSC - Identifying molecular features of cancers that predict response to anti-cancer drugs.
- http://www.cancerrxgene.org/

LINCS Cloud

- Interactive Browser and Apps for Managing the functional genomics data generated in the LINCS Project (Broad Institute)
- https://clue.io/

D3 – Data-Driven Documents

- D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

- https://d3js.org/

Interactive Visualization Tool

- https://www.bloomberg.com/billionaires/

Workflow

A workflow diagram depicts a series of actions that define a job or how work should be done. A workflow diagram visualizes how tasks will flow between resources, whether they're machines or people and what conditions allow the sequence to move forward. https://www.smartdraw.com/workflow-diagram/
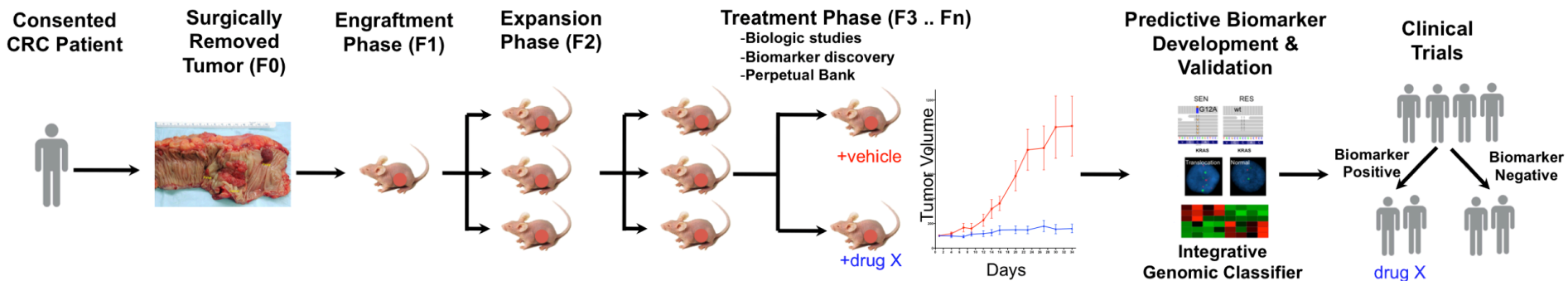
Workflow
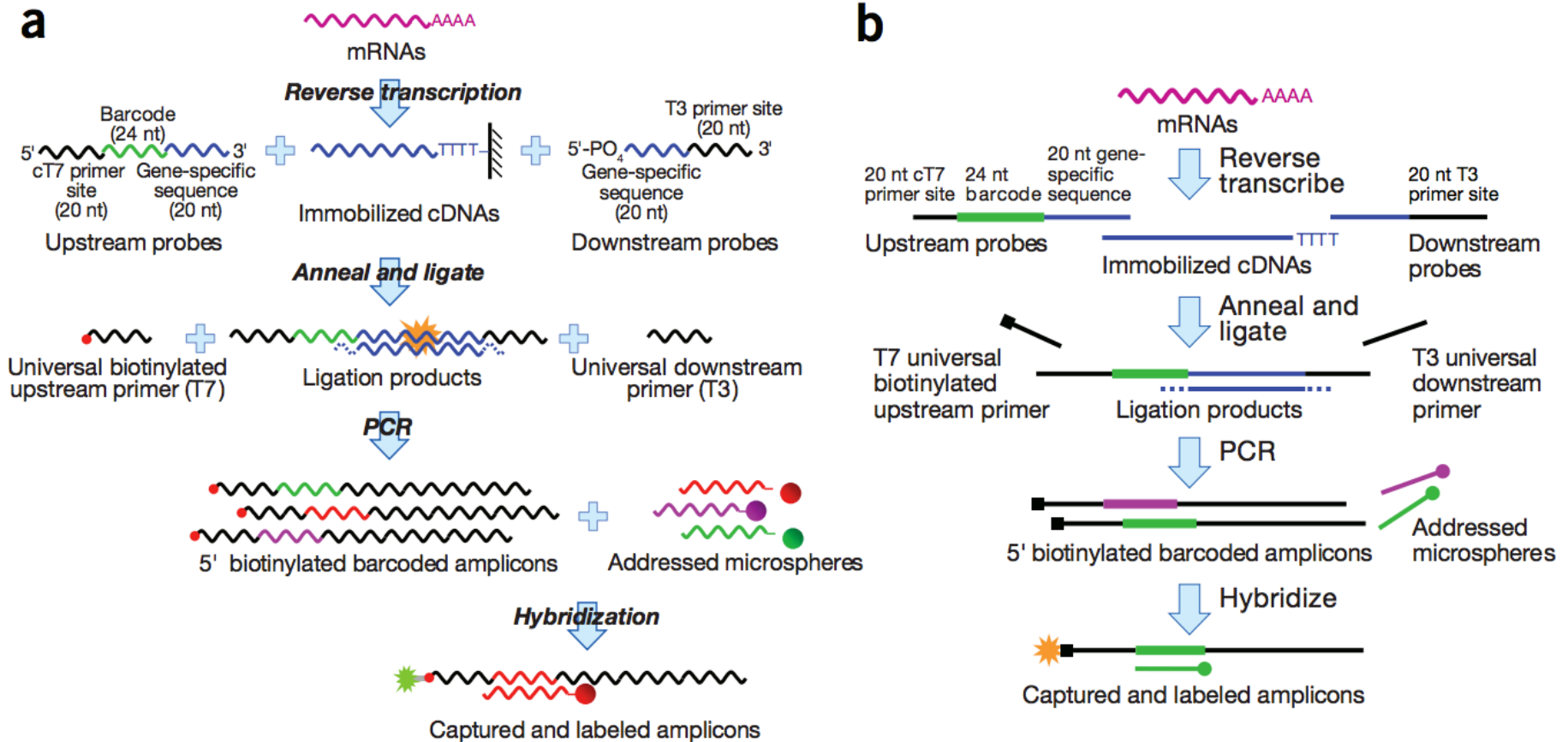
Figure 2 | Visual structure that matches the message. (**a**) Illustration showing a gene expression analysis technique. Reprinted from *Genome Biology*[4]. (**b**) The same elements organized according to the purpose of the illustration, which is to show a sequence of steps.
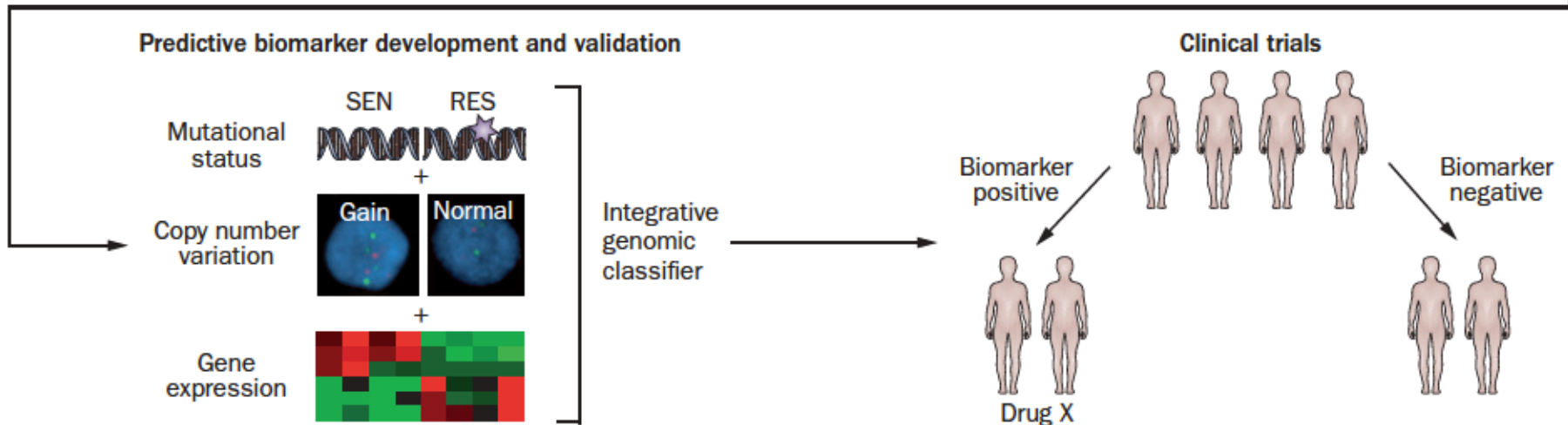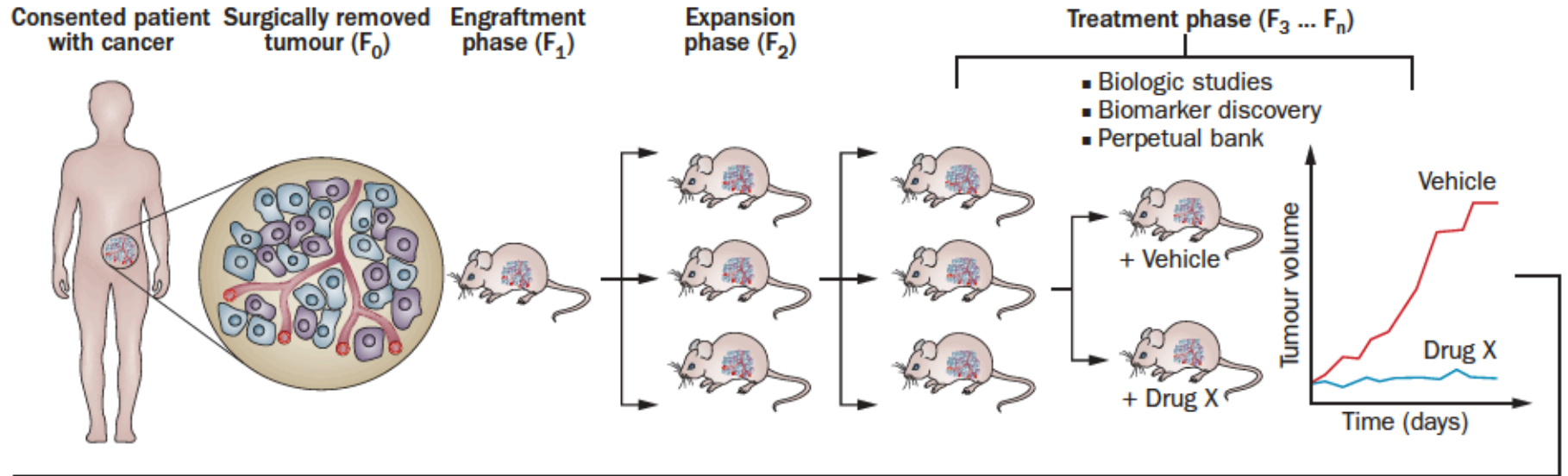
Wong B (2011) Nature Methods. http://www.nature.com/nmeth/journal/v8/n2/pdf/nmeth0211-101.pdf

Workflow

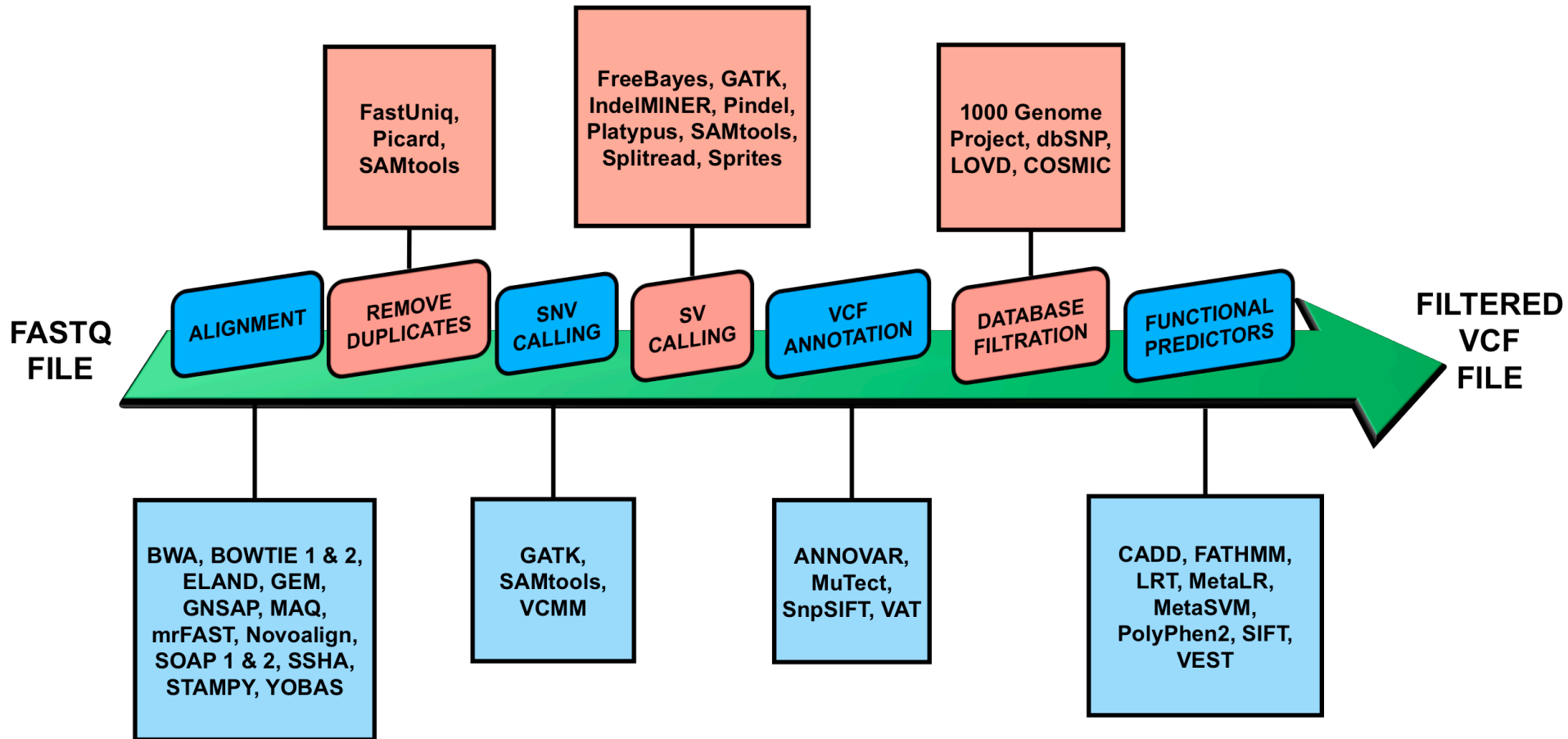## Original Design

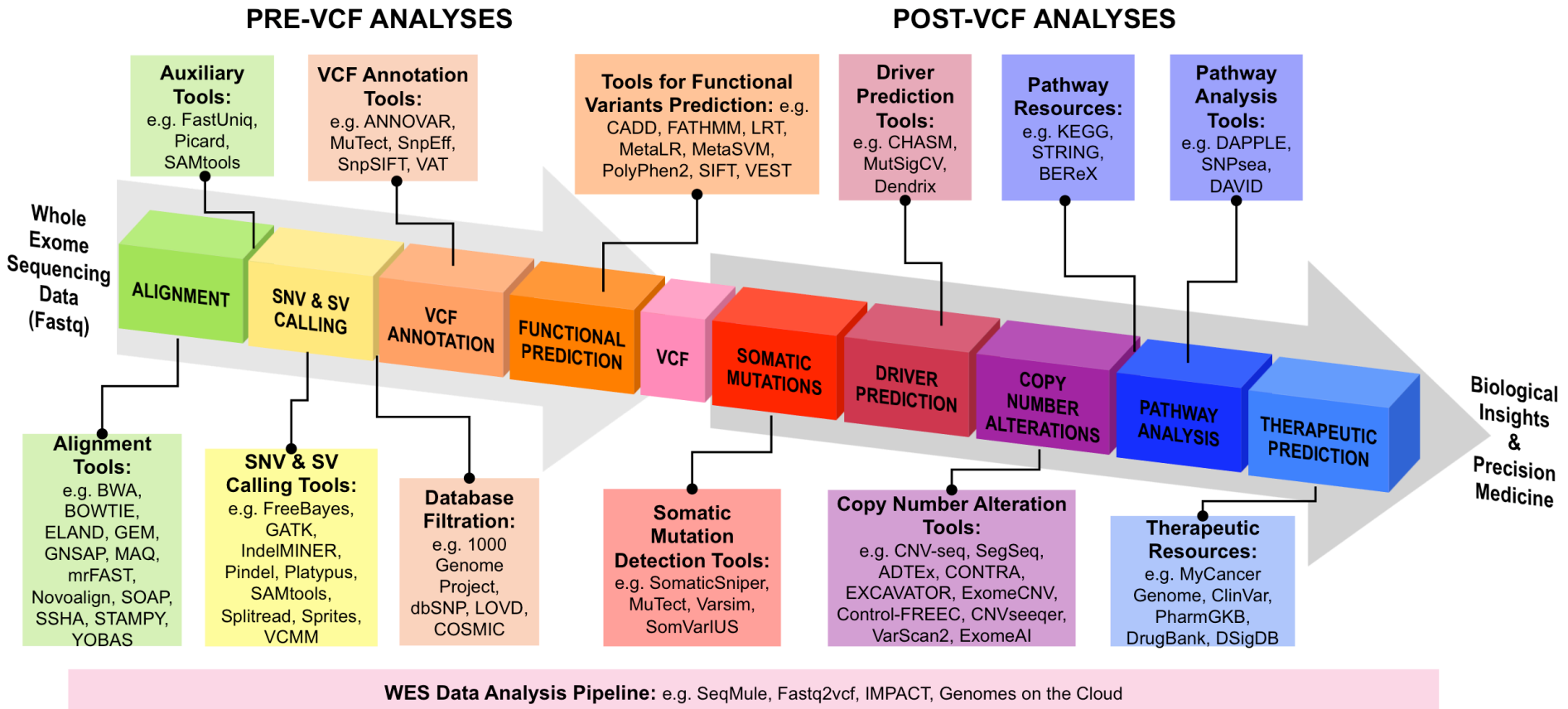# Workflow

(Tentler et al (2012). Nat Rev Clin Onc)

Final Design

Workflow

Original Design

Workflow

Final Design

More Resource

- VizBi – visualization of Biological Data Conference

- https://vizbi.org/