Recent developments in pharmaceutical R&D (a biostatistician's perspective)

Dhammika Amaratunga

OPA - May 2014

Pharmaceutical R&D

Identification and synthesis of new chemical entities suitable for therapeutic use.

2

Why is a statistician giving a pharma talk? (1)

"The best thing about being a statistician is that you get to play in everyone's backyard."

John W Tukey
(1915-2000)

3

Why is a statistician giving a pharma talk? (2)

"To understand God's thoughts we must study statistics, for these are the measure of His purpose."

Florence Nightingale
(1820-1910)

4

Why is a statistician giving a pharma talk? (3)

"All biology is becoming computational, much the same way it has became molecular."

Sydney Brenner
(1927-)

5

Drug Discovery: a bit of history

♦ Bark from the willow had been used as cure for various aches and pains since circa 1000 BC → acetylsalicylic acid in the 1800s → **aspirin**

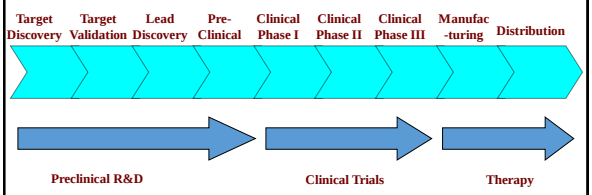
♦ Bark from the cinchona tree had been used in the Andes Mountains to cure fever → Europe in the 1600s to treat malaria → **quinine**

6

How are new drugs discovered?

Test Material	Test System
<19C	natural products	humans
19C-20C	+ chemicals (synthesis)	+ animals + organs + cells
Late 20C	+ combinatorial chemistry + structure-based chemistry + screening	+ proteins ("drug targets") + computer models

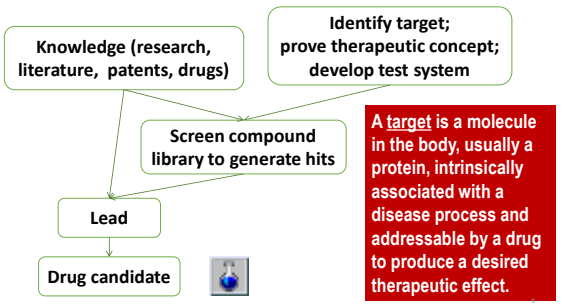
The pharma R&D process



Typically a drug candidate goes through as much as 10 years and $1billion of rigorous testing before it reaches the market … if it ever does … in fact, fewer than 1 in 10,000 drug candidates ever enters the market.

The Drug Discovery Process

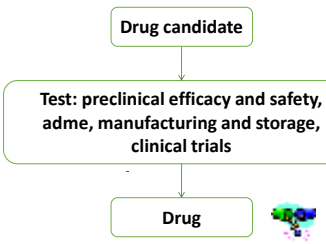
♦ **Purpose: Identify compounds that (selectively) address a protein associated with the disease process.**



A target is a molecule in the body, usually a protein, intrinsically associated with a disease process and addressable by a drug to produce a desired therapeutic effect.

The Drug Development Process

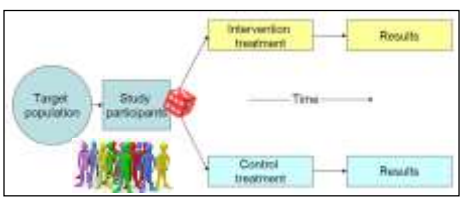
♦ **Purpose: Demonstrate that a drug candidate is efficacious, selective/safe, bioavailable, manufacturable, stable.**



A biomarker is a measurable indicator of a biological state; useful as an aid for decision making (e.g., diagnosis, prognosis, pharmacology, patient stratification, efficacy (surrogate).

Clinical trials

♦ Late-phase clinical trials are "hypothesis-driven" studies. They are designed to test the hypothesis that the treatment is more effective than the control.

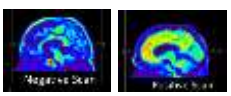


♦ Notes: *(1) Side effects are also tracked. (2) There are many variations in study design.*

Changing landscape of biology

♦ **The pharmaceutical R&D process is now being revolutionized by several technological developments …**

- Advances in genetics, genomics and proteomics
- Decoding of the human and other genomes
- Fast sequencing technologies
- Molecular imaging (CT, PET, MRI)
- Microelectronics/automation/nanotechnology



Plaque in Alzheimer's Disease

Impact of new technologies

♦ Diseases can be studied and characterized at a much deeper molecular level than has been possible so far.

♦ Multiple entities can be measured simultaneously (in high throughput) in "hypothesis-generating" studies.

♦ Enormous amounts of data are being collected. An emergent major challenge is to make the best use of this data.

Data → Information → Knowledge

13

Targets and biomarkers

♦ Two key possibilities are that there is now potential to develop drug targets and biomarkers that are much more direct and specific than current ones.

➢ *New drug targets*

➢ *New biomarkers*

14

Drug Targets

Target → **Drug**

Target = molecule in the body, usually a protein, intrinsically associated with a disease process and addressable by a drug to produce a desired therapeutic effect.

♦ Upto 2000 or so, only a few hundred drug targets (all single molecular) had been established.

➔ **Identifying a new drug target is like finding gold!**

Drug targets - example

♦ **Target: HMG-CoA reductase (HMGCR)** is a rate-controlling enzyme of the metabolic pathway that produces cholesterol.

♦ **Drug: Statins** lower cholesterol levels by inhibiting HMGCR; they are used to control cholesterol levels to reduce the risk of cardiovascular disease.

- Examples are rosuvastatin (Crestor), lovastatin (Mevacor), atorvastatin (Lipitor), pravastatin (Pravachol), simvastatin (Zocor), etc.

- Some red yeast rice (a Chinese herbal medicine) products contain a substance similar to lovastatin.

16

Biomarkers

♦ **Biomarkers** are useful in pharma R&D in a number of ways …

Biomarker = measurable indicator of a biological condition.

➢ *To characterize a disease*
- *diagnosing presence/absence of disease*
- *subtype/stage/severity/prognosis*

➢ *To assess post-treatment prognosis*
- *pharmacodynamics*
- *response to therapy*
- *surrogate marker*
- *patient selection/stratification*

Biomarker → **Personalized medicine**

17

Biomarkers – examples

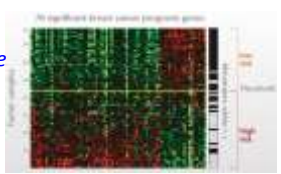
♦ Biomarkers may be univariate …

➢ *Glucose level in blood for diagnosis of diabetes*

➢ *HER2 elevation for treatment of HER2-positive breast cancer with Herceptin.*

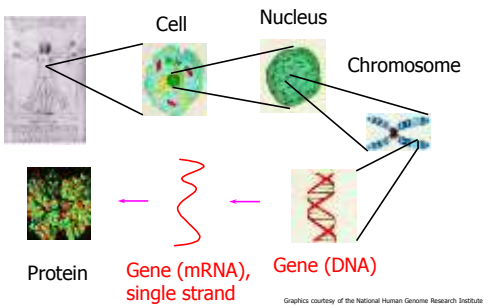
♦ … or multivariate …

➢ *MammaPrint is a 70-gene breast cancer gene signature which can be used to assess the risk that a breast tumor will metastasize.*



18

A very brief intro to molecular biology



Omics studies

♦ **The behavior pattern of genes/proteins can be studied …**

-- at the DNA level ("genetics")

To understand how variations in gene sequences are associated with specific diseases and/or with the body's response to a drug.

-- at the RNA level ("functional genomics")

To understand how variations in gene expression (RNA abundance levels) are associated with specific diseases and/or with the body's response to a drug.

-- at the protein level ("proteomics")

To understand how variations in protein levels (translational and post-translational) are associated with specific diseases and/or with the body's response to a drug.

-- at various other levels!

20

Data issues

♦ Within-study data:

- *Most experiments are run in high throughput mode with multiple features measured simultaneously in each sample.*
- *Preprocessing and analysis of data.*
- *Findings must be validated independently.*

♦ Across-study data:

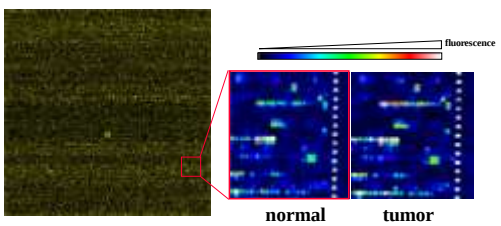
- *Data gathered from diverse sources (e.g., assays, pharmacokinetics, genomics, imaging)*
- *Store/organize/manage raw data*
- *Prepare/store/organize/manage derived data*
- *Data organization / management*

21

Within study data

♦ Many features (from a few hundreds to the millions) → **"high-dimensional"**

♦ Few samples (at most a few hundred)



Within study data - issues

♦ The good: Since "all" features are being studied, it is likely that key features and key feature combinations are present as part of the signal.

♦ The bad: Since "all" features are being studied, many carry no pertinent information and contribute to the noise.

♦ The challenge: **Separate the signal from the noise.**

♦ There is a strong potential for overfitting (incorrectly finding a random pattern very specific to this data, so findings cannot be reproduced on new data) – care must be taken if applying methods developed for "large *N*" problems (e.g., big data, data mining, machine learning).

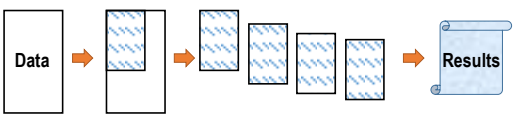
23

Methods for handling high-dimensional data

♦ **Regularization:** Use a model-fitting criterion which penalizes models with many non-zero parameters.

♦ **Filtering:** Reduce dimensionality (e.g., variable selection, principal component analysis).

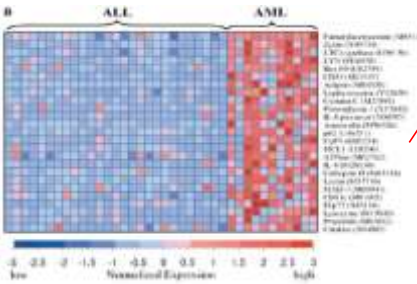
♦ **Ensemble:** Repeat filter+train many times (with perturbations) and aggregate the results (ensembles).



Note: Ensembles (1) work well (2) generate lists of useful features (3) provide performance assessments.

Early example

- Leukemia data (1999):
  - Study of ALL (acute lymphoblastic leukemia, *n*=27) vs AML (acute myeloid leukemia, *n*=11) using cDNA microarrays



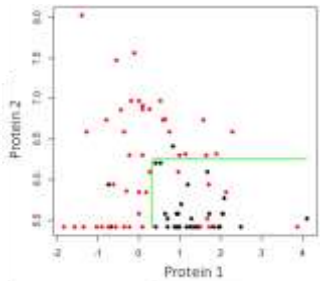
Gene list

Biological pathways and processes

25

Recent example

- Multiple myeloma data (2013):
  - Study of slow progressors (*n*=37) vs fast progressors (*n*=50)
  - Levels of 250+ proteins were profiled using the Myriad DiscoveryMAP® multiplexed immunoassay system



What role can Sri Lanka play?

- **Products**
  - *Natural product (herbal medicine)*
  - *Identification/synthesis of active component*
- **Demonstrate efficacy and safety**
  - *Strict market regulations in US/EU/JP*
  - *Clinical trial (GLP, GCP)*
  - *Epidemiological: retrospective (reproducibility)*
  - *Epidemiological: prospective*
- **Manufacturing considerations**
  - *Scale*
  - *Quality standards (GMP)*
- **Bioinformatics / Statistics**
  - *Interdisciplinary efforts!*

27

Wrap up

- **Book:**
  Amaratunga, Cabrera and Shkedy (2014): *Exploration and Analysis of DNA Microarray and Other High Dimensional Data* (second edition), Wiley.

- **Website (recent papers and software):**
  www. amaratunga.com
- **Email:**
  damaratung@yahoo.com

28