Using expressions Python

Regular Expressions

a|b	"a" or "b"
?	Zero or one
+	one or more
*	Zero or more
*?	Zero or more, but stop after 1 it matches
{N}	exactly N occurrences
{N,n}	from N to n occurrences

{,3} up to 3

[A-Z]	uppercase from "A" to "Z"
[0-9]	any number
[asdf]	any character in the list
[^asdf]	any but not a character in the list
.	any character, except newline

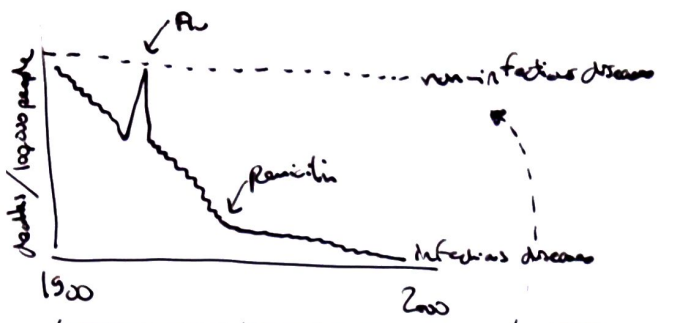
^	start of a line
$	end of a line
\s	whitespace
\S	non-whitespace
\w	word (alphanum + underscore
\W	non-word; i.e., special etc.

\d digit ~ [0-9]
\D non-digit
\w ~ [A-Za-z0-9_]

(...) capture group (3 chunks) any characters

```
import re
s="Biology = 5X fun!"
re.search("\d+", s)  # return if match
# <re.Match object; span=(6,7), match='5'>

# Search info genes/genomic coordinates
row = "5ox2__chr3__34548926__34551382__+"
re.search("\w+\s{2}chr\w{,2}\s{2}(\d+\s{2}){2}[\+\-]", row)
# also use/try the following regex, to minimize potential for unexpected matching
# \S+\s+chr\S+\s+\d+\s+\d+\s+[-+]
```



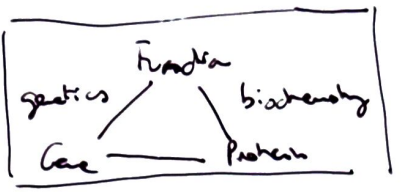
Independent of Molecular Medicine Knowledge

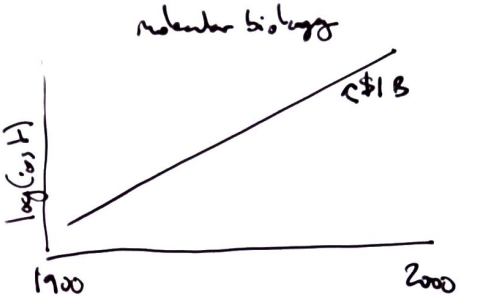
- serendipity (penicillin)
- herbal remedies (aspirin)
- hypothesis-driven (insulin)

~ natural products

Molecular biology

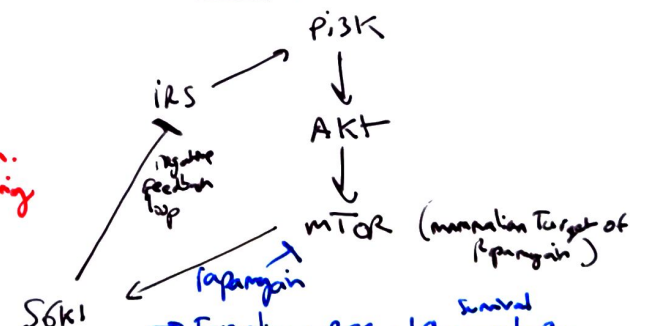
→ HIV: life cycle of the virus





Synthetic biology?

Cancer proliferation



⚠ rapamycin has no 100% activity → S6K1 ↓ → PI3K ↑

model: $K_1 \xrightarrow{E_1} K_1^* \xrightarrow{E_2} K_1$

↳ state of the system?

$$\frac{d[K_1]}{dt} = -k_1[K_1][E_1] + k_2[K_1^*][E_2]$$

approximation

Measurements of concentrations → establishment of model.
In the previous feedback loop system, inhibiting IRS, upstream of the cascade, gives a better control than rapamycin. And a combination is predicted to be even more efficient.

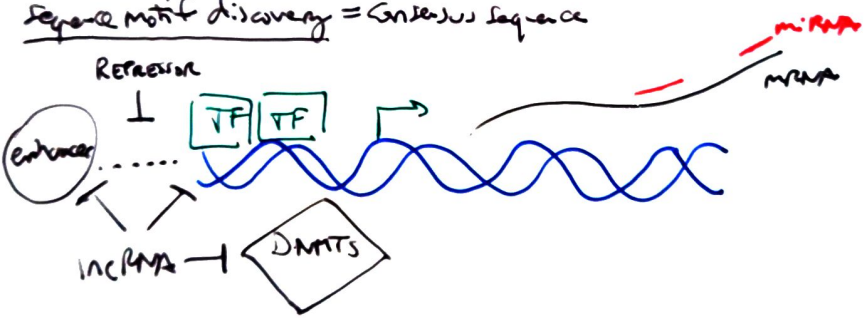
— specificity (target)
— therapeutic window

Genome Sequencing:

1995	H. influenza (bacteria)	1.8M
1997	S. Cerevisiae (yeast)	12M ← microarray
1998	C. elegans	100 M
1999	D. melanogaster	165M
2000	Human	3000M

Environmental Stress Response : ~1000 out of 6,000 genes in cerevisiae change expression regardless of the external stimulus

Gene Signature = set of genes useful for predicting something, e.g., disease.
↓
gene ontology database (GO) : knowledge on function, location, etc.
⚠ ~90% of datasets not annotated when perturbing a particular pathway
↳ transcription factors common to many gene expression
↑
Sequence motif discovery = Consensus sequence



easy to measure !

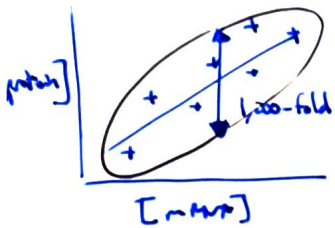
1999 : reverse genetics (yeast) : knock/perturbation
= modify genome → observe phenotype, systematically.

2000 : ChIP-chip → micro array Chromatin immunoprecipitation = protein-DNA interaction

(2009) Mikkelsen : ChIP-Seq

2001-2002 : protein-protein interaction / array

→ Mass spectrometry



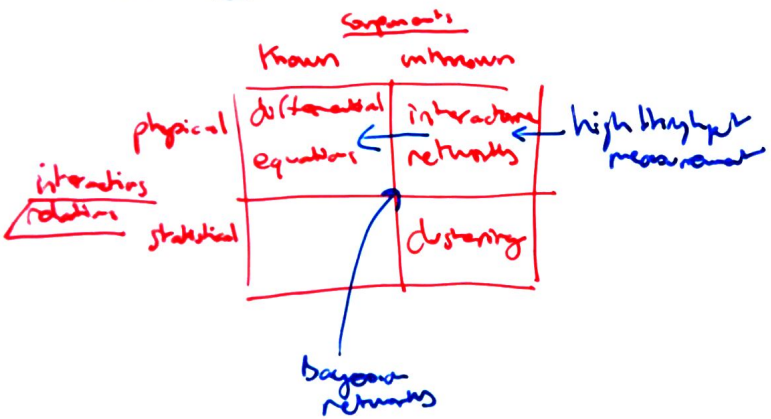
How much noise?

regulation → mRNA → protein

dogma true from a statistical level, but not very predictive

- reverse genetics, deletion/mutation, of one gene at a time → identify a gene that is a master regulator for the regulation of gene expression in response to a stimulus
- RNA-seq → identify set of genes / pathways dysregulated upon stimulus

Hits: RNA-Seq ∩ rev. genetics



EBs → mesoderm → Cardiac progenitors → Cardiomyocytes

d0 d4 d5.3 d10