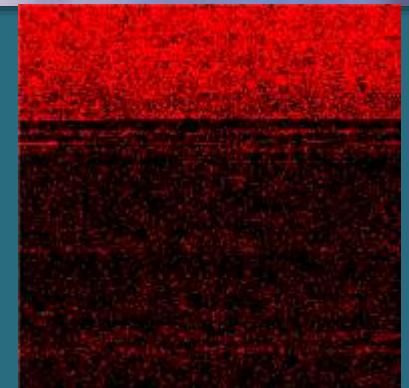


Cancer Type/Site Classification using Deep-Learning (Preliminary presentation slides)

S. Ravichandran, Ph.D
BIDS, FNLCR



Acknowledgements

- **NCI-DOE Pilot-1 Team**
- **BIDS**
 - Drs. George Zaki, Andrew Weissman, Mark Jensen and Eric Stahlberg
 - Amar Khalsa, Dr. Deb Hope, Anney Che, Hue Readron, Dr. Yongmei Zhao
 - Colleagues who reviewed the material

Feel free to follow-along

Github

- <https://github.com/ravichas/ML-TC1>

Introduction

- **This workshop is part of the NCI-DOE Pilot project knowledge/capability transfer efforts**
- **Goal is to share tools/techniques/solutions for cancer related problems. We often take a test-case and show how it works**
- **You would be able to take our test-case (code/scripts) and tune it to your needs**
- **We want to hear from you, please send us your feed-back**

Motivation: Cancer Prediction vs Cancer Detection

- **Cancer Prediction has been the major focus**
 - Prognosis, Recurrence, Susceptibility
- **Cancer Detection (classification of tumors/cancers) is lagging behind Prediction and we would like to share an application that might be useful**
 - Detect/Identify cancer type at an early stage

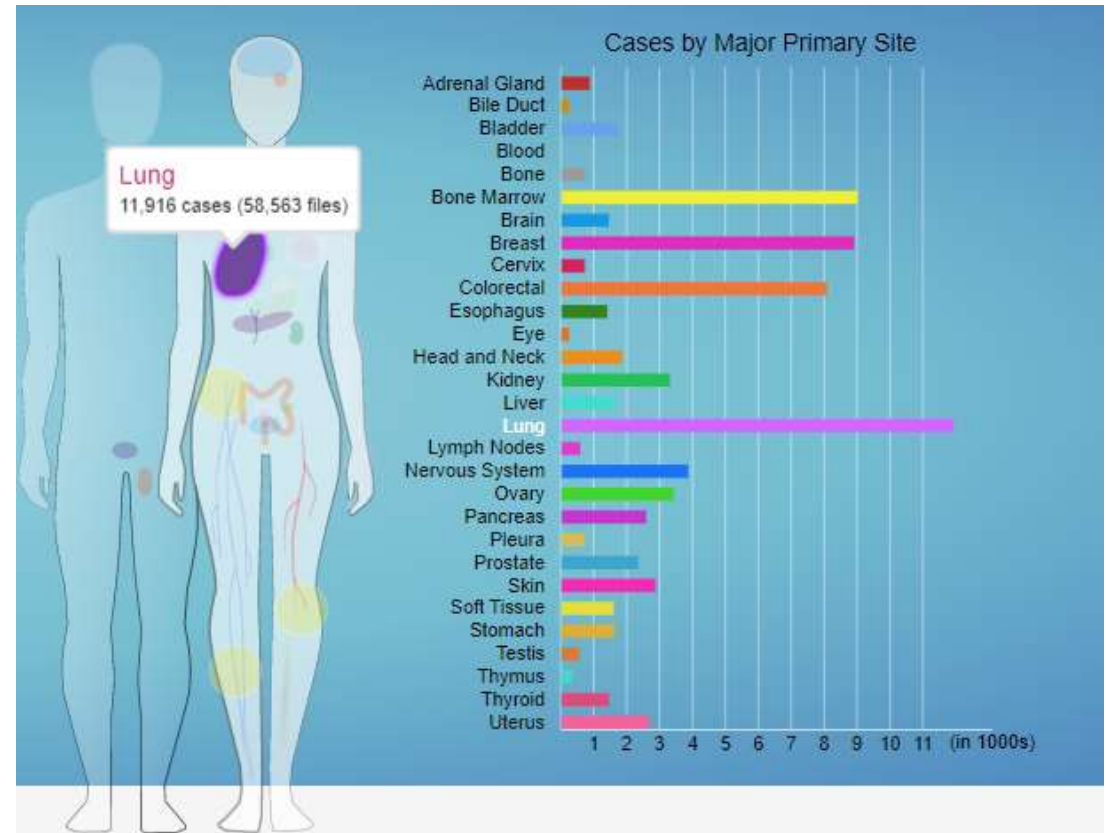
Goal(s)/Questions

- **Take genomic expression data from tumor/cancer samples and apply Deep-Learning to create cancer types/site(s) classifier models**
- **Are the expression profiles unique?**
- **Can we use the model as early cancer type detection**
 - Improving chance of early detection cure/survival?

Cancer Burden

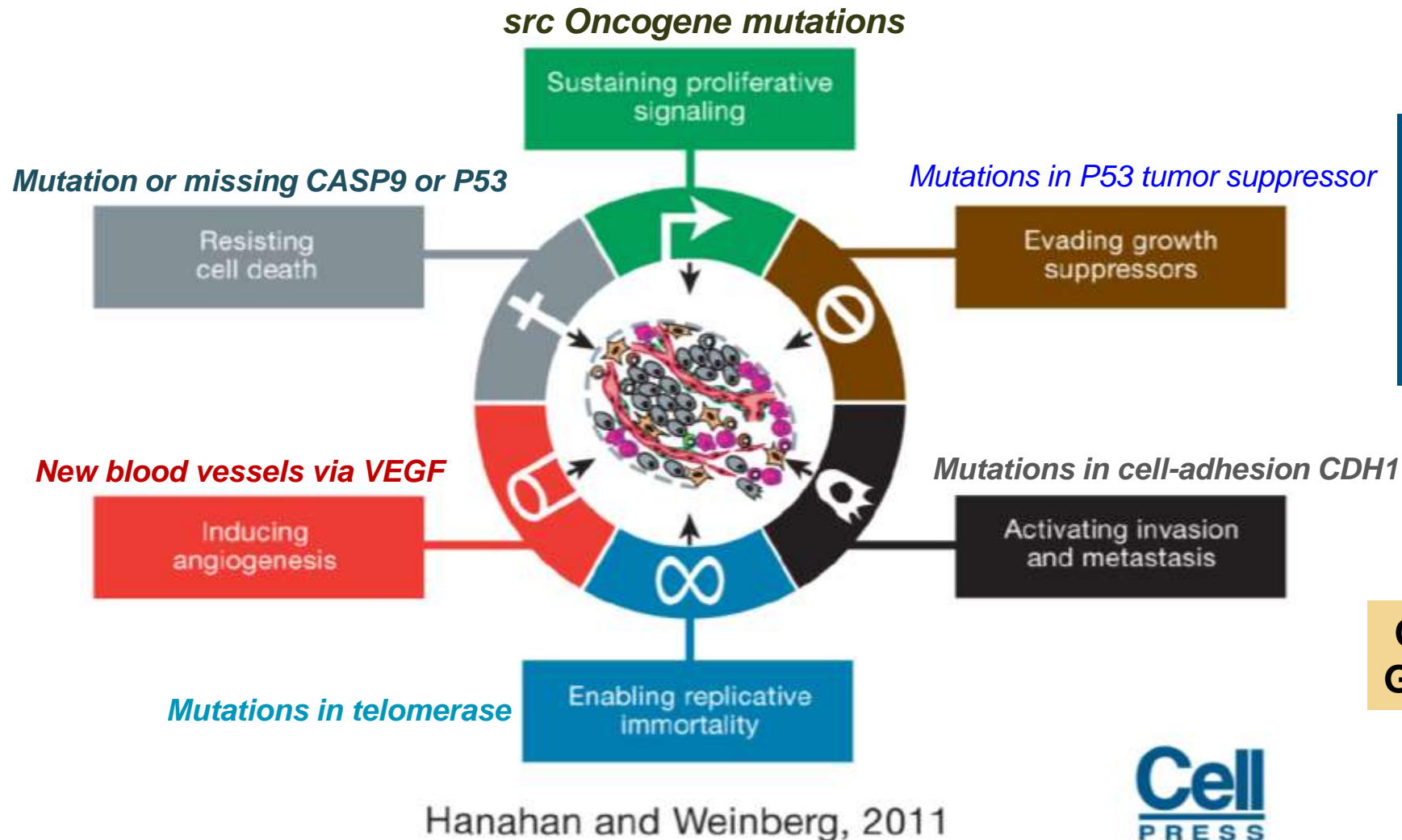
- **Cancer is a group of diseases with world-wide risk**
- **Acquired or somatic changes causes 90-95% of cancer (all types)**
 - *Source TCGA*
- **~ 200 forms of cancer**
 - *DOI: 10.5114/wo.2014.47136*
- **For 2020 in USA**
 - ~1.8M new cancer cases are expected
 - ~600K deaths will occur
 - *Source: American Cancer Society*

Figure from Genomic Data Commons



Hallmarks of cancer: Integral Components of Most Forms of Cancer

Hallmarks of Cancer: The Next Generation



REVIEW | VOLUME 100, ISSUE 1, P57-70, JANUARY 07, 2000

The Hallmarks of Cancer

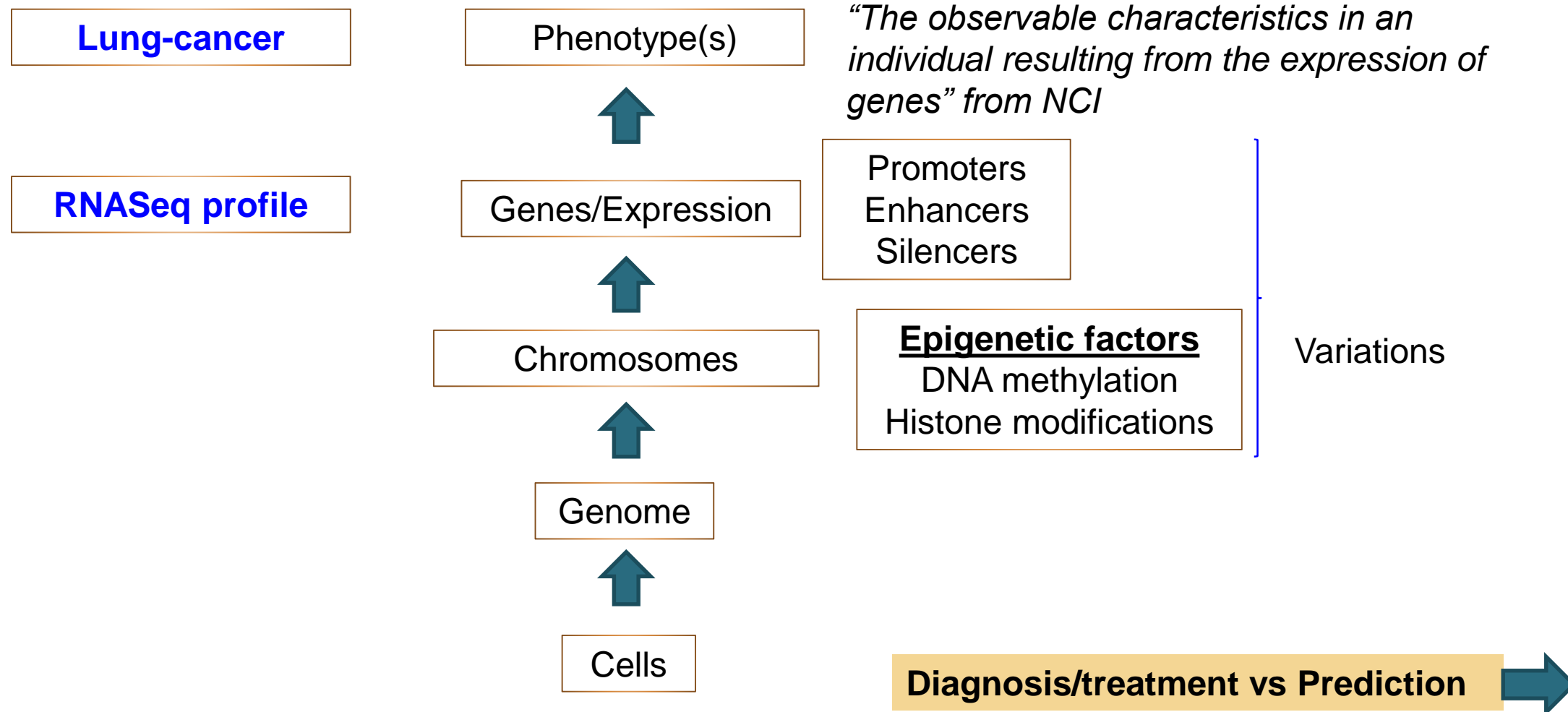
Douglas Hanahan • Robert A Weinberg

Open Archive • DOI: [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)

Overview of
Genotype/phenotypes?



Influence of genomic features on phenotypes: An overview



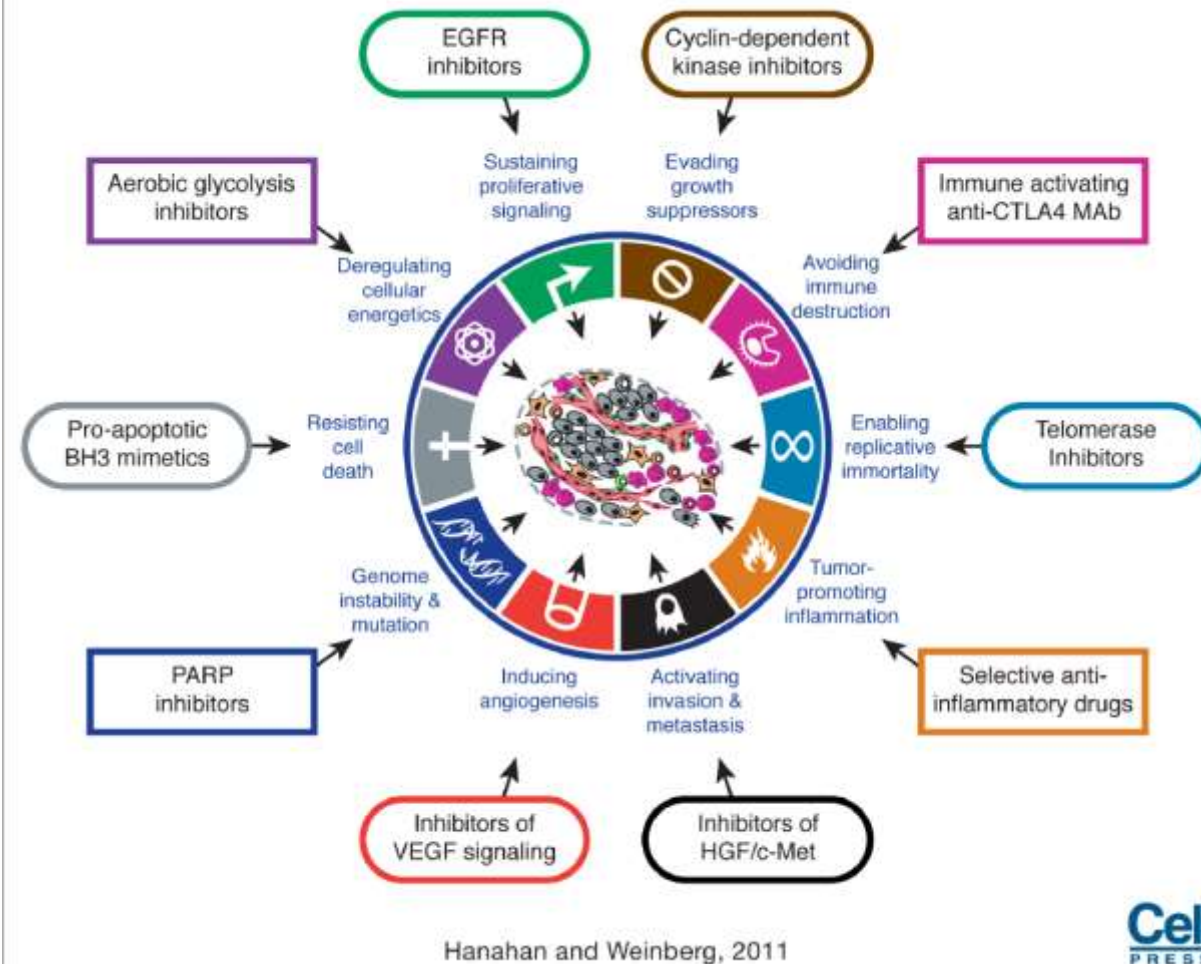
Treatment vs Type-Prediction

- **Treatment**

- Gene-centric (or a slice of pathway)
- Imatinib targeting BCR/KIT

- **Detecting Type**

- “The architecture of occurring genetic aberrations such as somatic mutations, CNVs, changed gene expression profiles, and different epigenetic alterations, is unique for each type of cancer.”, DOI: 10.5114/wo.2014.47136
- Complex & Multi-gene centric



Type-Prediction: Architecture of Genetic aberrations



The architecture of occurring genetic aberrations such as somatic mutations, CNV, changed gene expression profiles, and different epigenetic alterations, is unique for each type of cancer

DOI: [10.5114/wo.2014.47136](https://doi.org/10.5114/wo.2014.47136)

PERSPECTIVE

Understanding Genotype-Phenotype Effects in Cancer via Network Approaches

Yoo-Ah Kim, Dong-Yeon Cho, Teresa M. Przytycka*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America

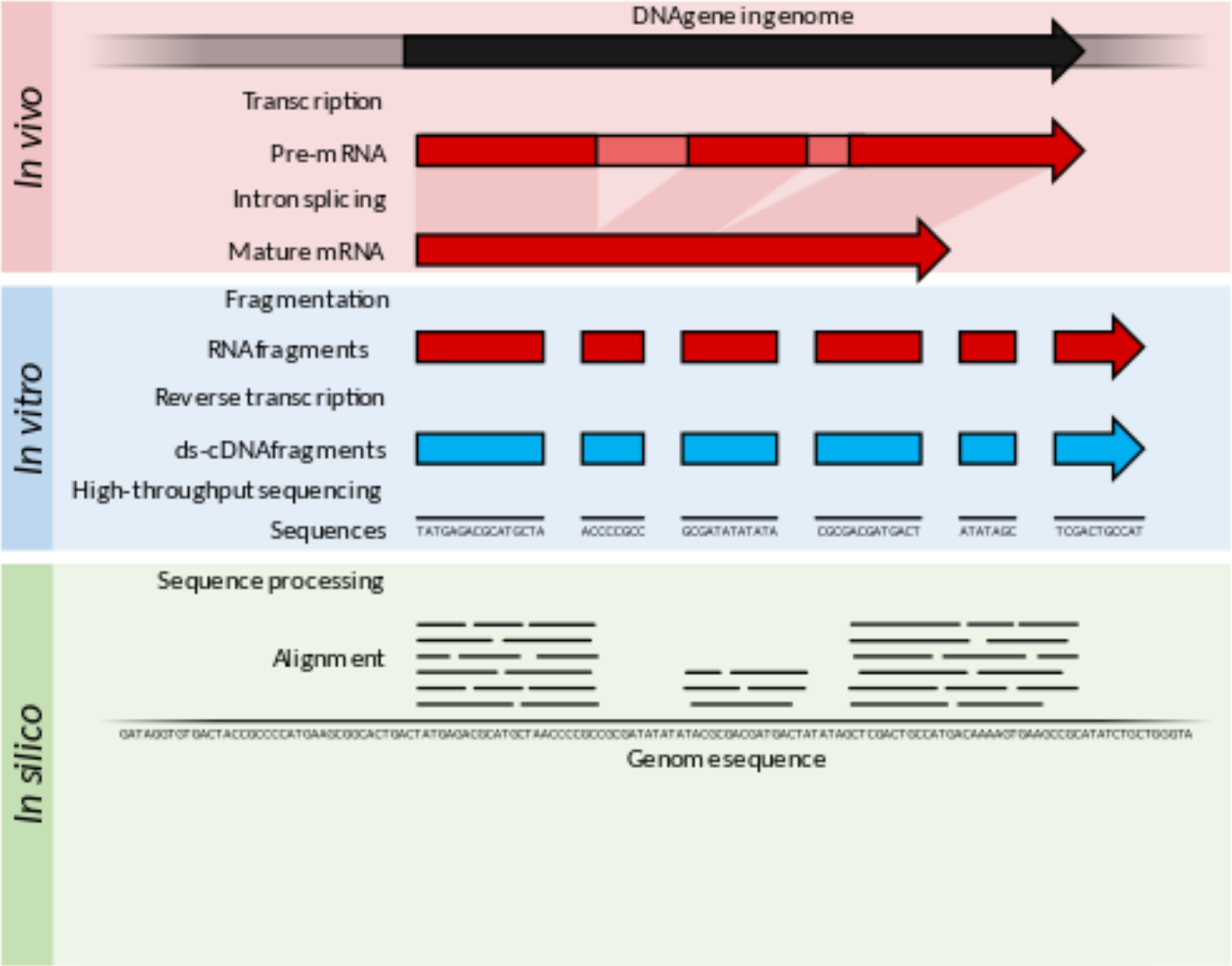
* przytyck@ncbi.nlm.nih.gov

Author Summary

Cancer is now increasingly studied from the perspective of dysregulated pathways, rather than as a disease resulting from mutations of individual genes. A pathway-centric view acknowledges the heterogeneity between genomic profiles from different cancer patients while assuming that the mutated genes are likely to belong to the same pathway and cause similar disease phenotypes. Indeed, network-centric approaches have proven to be helpful for finding genotypic causes of diseases, classifying disease subtypes, and identifying drug targets. In this review, we discuss how networks can be used to help understand patient-to-patient variations and how one can leverage this variability to elucidate interactions between cancer drivers.

Expression data

NGS



Spliced to become mature mRNA
mRNA is extracted

mRNA captured/fragmented/copied
into stable ds-cDNA
Sequenced

Reference Genome

NGS

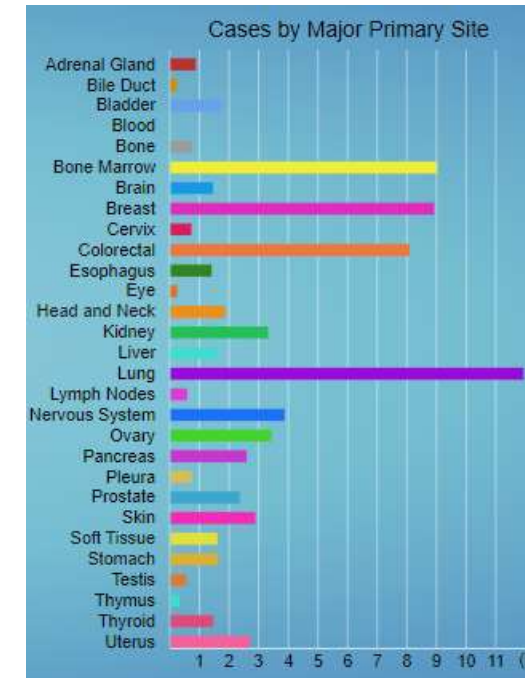
Data source: The Cancer Genome Atlas (TCGA)

- NIH launched TCGA Pilot Project – a public funded project
- Goal of creating a comprehensive “atlas” of cancer genomic profiles.
- Large cohorts of over 30 human tumors through large-scale genome sequencing and integrated multi-dimensional analyses.
- Contains Microarray and NGS data
 - RNASeq
 - miRNA seq
 - SNP based platforms
 -
- TCGA data is available via GDC

<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>

Data Harmonization: GDC (<https://gdc.cancer.gov/>)

- Data and metadata is submitted to the GDC in standard data types and file formats. Other data sources (Ex. TCGA) are also included
- Data are harmonized against a common reference genome (GRCh38)
- For this workshop, we will focus on TCGA Genomic expression data from GDC



Expression Data Quantification

- RC_g : Number of reads mapped to the gene
- RC_{g75} : The 75th percentile read count value for genes in the sample
- L : Length of the gene in base pairs; Calculated as the sum of all exons in a gene

$$FPKM-UQ = \frac{RC_g \times 10^9}{RC_{g75} \times L}$$

FASTQ

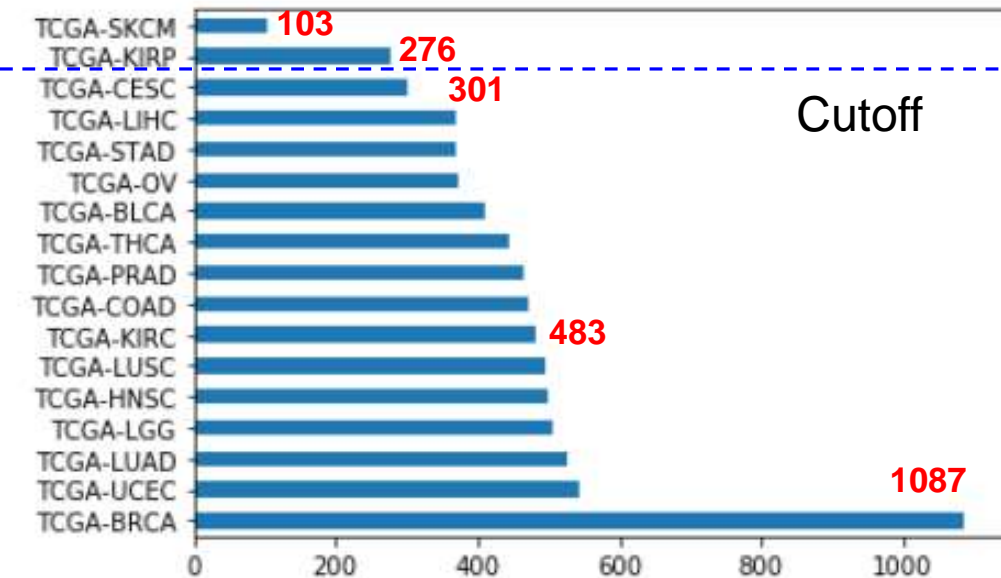
Alignment to Ref
Genome (SAM/BAM)

Quantification HTSeq

Gene Expression
(FPKM-UQ)

Fragments **P**er **K**ilobase of transcript per **M**illion mapped reads

How much data for modeling?



CODE	Cancer Site/Type
BRCA	Breast invasive carcinoma
UCEC	Uterine Corpus Endometrial Carcinoma
LUAD	Lung adenocarcinoma
LGG	Brain Lower Grade Glioma
HNSC	Head and Neck squamous cell carcinoma
LUHSC	Lung squamous cell carcinoma
KIRC	Kidney renal clear cell carcinoma
PRAD	Prostate adenocarcinoma
COAD	Colon adenocarcinoma
THCA	Thyroid carcinoma
BLCA	Bladder Urothelial Carcinoma
OV	Ovarian serous cystadenocarcinoma
STAD	Stomach adenocarcinoma
LIHC	Liver hepatocellular carcinoma
CEC	Cervical squamous cell carcinoma and endocervical adenocarcinoma

**300
samples
each**

Expression data from a sample

TCGA-BRCA

Genes	Expression
ENSG00000242268.2	1658.464179
ENSG00000270112.3	460.2343433
ENSG00000167578.15	52440.10096
ENSG00000273842.1	0
ENSG00000078237.5	68165.45626
ENSG00000146083.10	255959.2351
ENSG00000225275.4	0
ENSG00000158486.12	104.9473768
ENSG00000198242.12	4968556.658
ENSG00000259883.1	6108.999052
ENSG00000231981.3	0
ENSG00000269475.2	0
ENSG00000201788.1	0
ENSG00000134108.11	957330.2056
ENSG00000263089.1	3484.027373
ENSG00000172137.17	41485.9507
ENSG00000167700.7	226717.4208
ENSG00000234943.2	2082.245035
ENSG00000240423.1	310.5246749
ENSG00000060642.9	155863.9216
ENSG00000271616.1	0
ENSG00000234881.1	0
ENSG00000236040.1	394.4755669
ENSG00000231105.1	1583.312582
ENSG00000243044.1	0
ENSG00000182141.8	45538.60648
ENSG00000269416.4	119.0847054
ENSG00000264981.1	0

60,483
transcripts

Gene: AC090241.2 ENSG00000270112

Description

novel transcript, antisense to ST8SIA5

Location

[Chromosome 18: 46,756,487-46,802,449](#) forward strand.
 GRCh38:CM000680.2

About this gene

This gene has 8 transcripts ([splice variants](#))

Transcripts

Hide transcript table

Gene: DNAH3 ENSG00000158486

Description

dynein axonemal heavy chain 3 [Source:HGNC Symbol;Acc:[HGNC:2949](#)]

Gene Synonyms

DKFZp434N074, DLP3, Dnahc3b, Hsadhc3

Location

[Chromosome 16: 20,933,111-21,159,441](#) reverse strand.
 GRCh38:CM000678.2

About this gene

This gene has 6 transcripts ([splice variants](#)), [371 orthologues](#), [14 paralogues](#) and is a member of [1 Ensembl protein family](#).

Transcripts

Hide transcript table

Breast Cancer
60,484
transcripts

Data Preparation

Sample1

Sample2

Sample3

Sample4

Sample297

Sample298

Sample299

Sample300

Breast Cancer

Genes	Expression
ENSG00000242268.2	1655.4054179
ENSG00000270178.15	404.2341343
ENSG00000157513.15	52440.3300
ENSG0000013842.1	1384.21
ENSG00000173272.5	681.65.46626
ENSG00000146081.10	255699.2515
ENSG00000225275.4	0
ENSG00000184848.12	34794.97268
ENSG00000193842.12	4066556.58
ENSG0000025983.1	6108.990052
ENSG00000129813.1	0
ENSG00000139475.2	0
ENSG00000207881.1	0
ENSG00000134108.11	95730.2056
ENSG00000206898.1	438.027793
ENSG00000171917.12	107.1917
ENSG00000177008.7	2267.7420
ENSG00000234043.2	2082.24265
ENSG00000240921.5	103.5246749
ENSG00000126424.9	155083.9276
ENSG00000276561.1	0
ENSG00000234881.1	0
ENSG00000234044.1	394.317568
ENSG000001105.1	158.432584
ENSG00000248061.1	0
ENSG00000182141.8	455.38.0846
ENSG00000209410.4	131.0847054

Genes	Expression
ENSG0000024268.2	1658.464179
ENSG00000270112.3	406.231443
ENSG00000267815.3	52440.1300
ENSG00000273842.1	107.384211
ENSG00000278237.5	18556.9526
ENSG00000404083.10	25199.2515
ENSG0000025275.4	0
ENSG00000264842.12	194.347588
ENSG00000218422.12	4068556.58
ENSG00000209952.1	6108.999092
ENSG00000215813.0	0
ENSG00000203475.2	0
ENSG0000021788.1	0
ENSG00000114108.11	973130.2056
ENSG00000263089.1	8436.02773
ENSG0000027117.17	444.000000
ENSG00000267100.7	22617.4208
ENSG00000234943.2	2082.24035
ENSG00000240423.1	103.5246749
ENSG00000240423.1	155063.9276
ENSG00000275616.1	0
ENSG00000234885.1	0
ENSG00000204610.1	394.475837
ENSG0000021105.1	1828.31580
ENSG00000243044.1	0
ENSG00000218214.8	4533.676048
ENSG00000269416.4	139.084754

Genes	Ensembl	Exonisation
ENSG00000242628.2	3058	484179
ENSG00000210212.3	40	234343
ENSG00000207726.5	52440	35000
ENSG00000217842.1	1	7842.1
ENSG00000278273.5	68355	46656
ENSG00000408110.8	205999	253
ENSG00000252574.0	1	52574
ENSG00000254865.2	1	54973268
ENSG00000182462.12	4086556	58
ENSG00000259833.1	6108	899052
ENSG00000210811.0	1	10811
ENSG00000209475.2	1	9475
ENSG00000210788.1	1	10788
ENSG00000134108.11	957330	2056
ENSG00000088918	3484	673773
ENSG00000271217.17	1	71217
ENSG00000167600.7	22617	4308
ENSG00000234433.2	2082	249525
ENSG00000240423.1	310	5246740
ENSG00000210572.9	3058	9276
ENSG00000276561.0	1	76561
ENSG00000234811.1	1	34811
ENSG00000240404.1	394	473563
ENSG00000211055.1	1583	921582
ENSG00000243041.1	1	43041
ENSG00000182141.8	45538	38408
ENSG00000249164.6	118	684705

Genes	Expression
ENSG00000242268.2	1658.464719
ENSG00000271713.1	460.234343
ENSG00000175918.5	52440.130031
ENSG00000134431.1	244.138411
ENSG00000178277.5	6838.456266
ENSG00000146081.0	25059.95125
ENSG00000252574.4	6.108
ENSG00000174218.1	947.947588
ENSG00000182944.2	4096856.6818
ENSG00000259831.1	6.108
ENSG00000159813.1	6.108
ENSG00000130475.2	6.108
ENSG00000178811.0	6.108
ENSG00000134108.11	957330.2056
ENSG00000160839.1	4848.027737
ENSG00000171717.1	17.171717
ENSG00000177007.7	226717.4048
ENSG00000249434.2	2082.245055
ENSG00000249243.1	153.5246749
ENSG00000151051.9	150883.8797
ENSG00000274656.1	6.108
ENSG00000148811.0	6.108
ENSG00000246041.1	394.475669
ENSG00000141051.9	1538.312683
ENSG00000239044.1	6.108
ENSG00000182148.8	45538.06468
ENSG00000239416.4	138.084704

Genes	Time	Expression
ENSG00000242688.2	12658.2	1568.464179
ENSG00000270123.1	60	204.234433
ENSG00000271719.1	18	52440.10096
ENSG00000271842.1	18	184.234433
ENSG00000278275.1	68	68165.46626
ENSG00000280148.1	20	25599.2351
ENSG00000252574.0	6	0
ENSG00000254572.1	18	194.945778
ENSG00000280426.12	12	408955.65
ENSG00000259831.1	68	6196.99052
ENSG00000251983.1	0	0
ENSG00000254075.2	20	20.235473
ENSG00000217881.1	0	0
ENSG00000214108.11	15	95730.2056
ENSG00000280898.1	18	3484.657173
ENSG00000271217.1	18	174.235473
ENSG00000257907.1	70	22517.42048
ENSG00000240443.1	20	2082.26525
ENSG00000244942.1	130	52467.49
ENSG00000246429.9	15	55683.5216
ENSG00000276561.1	6	0
ENSG00000248811.1	0	0
ENSG00000246041.1	104	475.375567
ENSG0000021155.1	18	158.464179
ENSG00000230404.1	38	33.425473
ENSG00000292148.8	4538	405.468
ENSG00000281644.8	119	88704.54

Genes	Log2 Expression
ENSG0000024268.2	1658.464179
ENSG00000270113	460.2934431
ENSG0000027812.3	52440.10096
ENSG0000027821.5	68265.4566
ENSG0000014800.1	20599.91281
ENSG0000025274.7	0
ENSG0000025866.1	52.9472768
ENSG0000028412.2	498826.586
ENSG0000025883.1	6308.999052
ENSG0000023198.3	0
ENSG0000028472.9	0
ENSG0000001780.1	0
ENSG0000013408.1	957330.2056
ENSG0000001008.1	3484.027173
ENSG0000017211.7	147.71171
ENSG0000016700.2	22673.4078
ENSG0000024943.2	2082.240525
ENSG0000024531.3	310.5246749
ENSG0000026624.9	550683.9216
ENSG0000027456.1	0
ENSG0000023488.1	0
ENSG0000023040.1	394.4755478
ENSG0000011026.1	833.312582
ENSG0000004304.1	0
ENSG0000018214.8	45338.60648
ENSG0000004945.6	138.087034

Gene	Exp	Expression
ENSG00000242088.2	1658	4641.79
ENSG00000270112.3	460	23443.03
ENSG00000257816.3	1044	10096
ENSG00000246412.3	5	1042
ENSG00000248235.5	68165	45626
ENSG00000260730.3	205919	2531
ENSG00000252574.4	0	0
ENSG00000254686.12	12	9877.58
ENSG00000258042.2	194	49955.68
ENSG00000258083.1	6388	69002.952
ENSG00000251983.1	0	0
ENSG00000250425.2	0	0
ENSG00000201788.11	1408	11
ENSG00000213001.18	95730	2056
ENSG00000203089.18	3468	62773
ENSG00000217117.12	48	9502
ENSG00000256700.7	22671	2420
ENSG00000248493.2	2082	2405
ENSG00000240523.1	531	52467.99
ENSG00000240626.2	19	52662
ENSG00000272565.1	0	0
ENSG00000234881.1	0	0
ENSG00000236040.1	39	45563.69
ENSG00000211815.1	8383	31262
ENSG00000243441.6	40338	60548
ENSG00000249446.4	128	48705.4

Gene	Expression
ENSG00000244208.2	1658.464179
ENSG000002701.123	460.234343
ENSG000002785.23	52404.30096
ENSG000002784.2	178.000000
ENSG000002786.13	68165.46526
ENSG000002400.150	25599.2511
ENSG000002252.574	50
ENSG000002248.126	184686.12
ENSG000002582.122	194.947378
ENSG000002582.122	408955.68
ENSG000002598.11	6108.99052
ENSG000002319.13	50
ENSG000002345.2	50
ENSG000002370.188	1
ENSG000002143.018	957330.266
ENSG000002008.11	3484.027173
ENSG000002721.17	178.000000
ENSG000002677.000	22817.4408
ENSG000002349.42	2082.45025
ENSG000002403.1	310.5246749
ENSG000002624.2	25863.12
ENSG000002715.261	50
ENSG000002348.11	50
ENSG000002340.041	394.475678
ENSG000002311.25	833.112562
ENSG000002444.1	50
ENSG000002321.418	45338.40548
ENSG000002494.16	119.084704

Lung Cancer

Genes	Expression
ENSG00000204268.2	3656.464179
ENSG00000207178.1	406.2349433
ENSG00000105753.5	52440.130046
ENSG00000138412.1	1884.611381
ENSG00000273175.7	68165.464626
ENSG00000146088.10	25599.93125
ENSG0000025274.0	0
ENSG00000105753.5	5947.947768
ENSG00000198422.12	4066556.58
ENSG00000258383.1	6108.990052
ENSG00000259883.1	0
ENSG00000258475.2	0
ENSG00000257888.1	0
ENSG00000134108.11	95730.2056
ENSG00000260389.1	3484.027734
ENSG00000105753.5	4887.7907
ENSG00000270075.9	22617.4208
ENSG00000234035.2	2082.244505
ENSG00000140423.1	130.5248749
ENSG00000105753.5	155863.9216
ENSG00000275616.1	0
ENSG00000234881.1	0
ENSG00000258475.2	394.473562
ENSG00000111058.1	358.3125560
ENSG00000240044.1	0
ENSG00000182141.8	455.3846058
ENSG00000269416.4	118.0847554

[illegible]

Genes	Fermentation
ENSG00000242682.2	3528.484719
ENSG00000210713.15	404.234343
ENSG00000207578.15	52440.13000
ENSG00000208021.0	139.0121
ENSG00000236775.7	68355.46624
ENSG00000140081.10	20599.2511
ENSG00000226754.0	0
ENSG00000214485.12	128.947368
ENSG00000210842.12	400655.56
ENSG00000259833.1	618.99052
ENSG00000231983.0	0
ENSG00000259832.2	0
ENSG00000237881.0	0
ENSG00000134028.11	957330.2056
ENSG00000203809.1	3484.037373
ENSG00000272137.17	4445.855
ENSG00000167902.17	228717.2306
ENSG00000214943.2	2082.140253
ENSG00000240423.11	130.5248749
ENSG00000211919.19	155363.9276
ENSG00000274561.0	0
ENSG00000234881.1	0
ENSG00000230404.1	3304.475569
ENSG00000208041.15	583.32582
ENSG0000021214.8	45338.28054
ENSG00000209416.18	118.087454

ENS	Games	Perceptions
ENS0000000242082	3526	484179
ENS0000000217131	460	234343
ENS0000000750755	5240	10096
ENS0000000178421	184	9142
ENS0000000378275	68365	46626
ENS0000000418000	25059	2351
ENS0000000252754	0	0
ENS0000000248452	394	947768
ENS0000000398422	408656	68
ENS0000000298831	6108	999052
ENS0000000219883	0	0
ENS0000000298422	0	0
ENS0000000278881	0	0
ENS000000031408.11	95730	2056
ENS0000000280891	3488	67377
ENS0000000721372	44	9273
ENS0000000278027	22973	4080
ENS0000000240424	2082	40525
ENS0000000404231	130	526479
ENS0000000298542	15383	8726
ENS0000000716161	0	0
ENS0000000214881.1	0	0
ENS0000000360401	394	475569
ENS0000000235052	5383	33264
ENS0000000240044	0	0
ENS0000000218214	45338	60548
ENS0000000294166	118	587054

[illegible]

Genotype	Expression
ENTG0000024268.2	1256.464179
ENTG0000027173.18	540.234343
ENTG000002605.2	12420.10096
ENTG0000027173.18	0
ENTG0000027217.5	6835.4456
ENTG0000024608.10	21599.91281
ENTG0000025274.6	0
ENTG0000024688.13	14.9673738
ENTG0000024942.12	49865.56
ENTG0000025983.1	6108.99252
ENTG0000023198.10	0
ENTG0000025672.2	0
ENTG0000021788.1	0
ENTG0000023408.11	95730.266
ENTG0000023008.1	3484.02733
ENTG0000027217.17	414.485705
ENTG0000027900.3	22571.4708
ENTG0000024943.2	2082.24051
ENTG0000024042.10	120.5246749
ENTG0000024112.9	15186.12156
ENTG0000027165.1	0
ENTG0000024881.1	0
ENTG0000023040.1	394.475563
ENTG0000023185.1	583.312582
ENTG0000023004.1	0
ENTG0000028241.8	45738.50648
ENTG0000026546.6	119.087054

EN	Exposition
ENSG000000042568.2	1658.464199
ENSG00000007012.3	460.234343
ENSG000000075818.5	25404.10096
ENSG00000006237.5	68165.46656
ENSG00000008280.3	205919.2181
ENSG000000025275.4	0
ENSG000000068168.12	94.9473768
ENSG00000008244.2	408965.6858
ENSG00000005983.1	6308.990052
ENSG000000021983.3	0
ENSG00000006452.1	0
ENSG000000017088.1	0
ENSG000000034028.11	95730.2656
ENSG000000063089.1	3844.02773
ENSG000000072145.1	485.9507
ENSG000000007700.7	23571.0
ENSG00000004493.2	2082.2405
ENSG00000004023.3	130.524679
ENSG00000004493.2	53863.9212
ENSG000000075166.1	0
ENSG00000004881.1	0
ENSG000000026040.1	394.475647
ENSG000000038519.1	383.117582
ENSG00000003044.1	0
ENSG000000082416.8	45338.0648
ENSG000000069456.4	119.087054

ENSG00000242568	Expression
ENSG00000242568	1658.46479
ENSG00000270112	462.234343
ENSG00000268213	2504.10096
ENSG00000276375	68365.46656
ENSG00000268103	259599.2151
ENSG00000252754	0
ENSG00000256866	124.9473738
ENSG00000258042	498985.56
ENSG00000259883	6198.99052
ENSG00000231983	3
ENSG00000259672	0
ENSG00000201788	1
ENSG00000234028	957330.256
ENSG00000238011	3484.02773
ENSG00000272707	14485.87
ENSG00000249007	23971.4708
ENSG00000234423	2082.2405
ENSG00000240493	130.5246749
ENSG00000259443	5186.912
ENSG00000275364	1
ENSG00000234881	0
ENSG00000236040	394.45765
ENSG00000243003	383.312582
ENSG00000243044	1
ENSG00000282148	45338.60648
ENSG00000269446	129.087054

Kidney Cancer

Genes	Expression
ENSG00000242626.2	1658.464179
ENSG00000210712.3	404.234343
ENSG00000245258.15	52440.10096
ENSG00000273842.1	0
ENSG00000244442.12	6835.46526
ENSG00000140831.10	250959.2351
ENSG00000225878.1	0
ENSG00000144544.10	104.947178
ENSG00000244442.12	68465.6168
ENSG00000259881.1	0.006199052
ENSG00000251981.3	0
ENSG00000249475.2	0
ENSG000002788.1	0
ENSG00000134081.11	973730.2056
ENSG00000203893.1	3484.027371
ENSG00000171317.17	44485.9507
ENSG0000017608.7	22817.4238
ENSG00000244442.12	2982.740105
ENSG00000240421.13	5120.52649
ENSG0000020682.9	151863.9216
ENSG00000274861.1	0
ENSG00000244442.12	0
ENSG00000230401.1	394.4755669
ENSG00000231105.1	1581.312582
ENSG00000240441.1	0
ENSG00000242144.18	455.38458
ENSG00000244442.12	19.082075

Genes	Expression
ENSG00000242058.2	1655.464179
ENSG00000270112.3	406.234343
ENSG0000006578.15	52440.10000
ENSG00000278942.1	0
ENSG00000278927.5	18581.45262
ENSG00000046083.10	255959.25131
ENSG00000252275.4	0
ENSG00000158496.12	104.9471708
ENSG00000252075.2	1656.658.658
ENSG00000259883.1	6108.999052
ENSG00000231881.3	0
ENSG00000260945.2	0
ENSG00000252075.2	1656.658.658
ENSG00000344028.11	957330.2056
ENSG00000263073.1	34848.023891
ENSG00000172137.17	14185.9507
ENSG00000046703.1	2267.142670
ENSG00000270000.1	24.24033
ENSG00000240042.1	1030.524649
ENSG00000259126.9	155861.9216
ENSG00000272816.1	0
ENSG00000260482.1	0
ENSG00000261600.1	394.4755669
ENSG00000231105.1	1583.312582
ENSG00000340344.1	0
ENSG00000282149.1	458.458458
ENSG00000252075.2	10.0842054

Genes	Expression
ENSG00000242682	2658.464179
ENSG00000210112	460.293443
ENSG000001675815	5240.2000
ENSG00000178482	1
ENSG00000232175	1658.45426
ENSG00000140081	20599.9121
ENSG00000225754	1
ENSG00000158412	134.947768
ENSG00000140312	408.056
ENSG00000159811	6108.9902
ENSG00000131981	3
ENSG00000169475	0
ENSG00000200781	1
ENSG00000134028	197330.2506
ENSG00000163089	3484.027031
ENSG00000172117	41485.9507
ENSG00000167007	2287.614208
ENSG00000140041	292.294505
ENSG00000140311	130.524679
ENSG00000060442	151861.9216
ENSG00000171661	0
ENSG00000140419	1
ENSG00000236040	394.4755669
ENSG00000131052	13182.11025
ENSG00000140044	1
ENSG00000182148	455.38.30548
ENSG00000198448	1

Gene	Expression
ENSG0000024268.2	1658.464179
ENSG00000170112.3	460.234343
ENSG00000165758.15	53440.3006
ENSG00000173862.1	0
ENSG00000172875.2	68385.4626
ENSG00000146083.10	25099.971
ENSG00000152976.4	0
ENSG000015446.12	130.947378
ENSG00000142312.2	104976.658
ENSG00000159811.1	6308.999052
ENSG00000139183.3	0
ENSG00000164075.1	0
ENSG00000149828.2	0
ENSG00000134118.10	95730.2056
ENSG00000163089.1	3484.02705
ENSG00000172137.17	14485.957
ENSG00000146072.1	22671.42470
ENSG00000146073.1	24505
ENSG00000140031.3	130.536749
ENSG0000060642.9	15581.9152
ENSG00000171616.1	0
ENSG00000134891.1	0
ENSG00000130401.1	394.455669
ENSG00000131055.1	1833.31282
ENSG00000182144.1	453.584028
ENSG00000142948.4	19.982676

Genes	Expression
ENSG00000242268.2	1658.484179
ENSG00000270172.1	404.2343433
ENSG00000157816.5	52440.10096
ENSG0000073842.1	0
ENSG0000027382.5	68165.466256
ENSG00000140833.0	205599.2351
ENSG00000157816.5	52440.10096
ENSG00000158486.12	104.93417978
ENSG00000122412.2	490856.558
ENSG00000219883.1	618.999052
ENSG00000211883.3	0
ENSG00000208475.2	0
ENSG00000217881.0	0
ENSG00000217881.0	0
ENSG00000217881.0	0
ENSG00000230810.1	3648.027953
ENSG00000217327.7	41485.9507
ENSG0000017700.7	228737.4028
ENSG00000217327.7	41485.9507
ENSG00000240623.1	5130.526649
ENSG00000206042.9	150883.2616
ENSG00000271683.1	0
ENSG00000214883.1	0
ENSG00000230401.1	394.475660
ENSG00000211109.1	5138.312582
ENSG00000212414.8	45338.0646
ENSG00000204265.4	119.0887054

Genes	Expression
ENSG0000024268.2	3058.404179
ENSG00000070012.3	460.2349433
ENSG00000167578.15	52440.130043
ENSG00000273842.1	0
ENSG00000078237.5	48165.45626
ENSG00000140033.10	255959.2513
ENSG00000167578.15	52440.130043
ENSG00000158486.12	304.9471268
ENSG00000182446.12	4896556.15528
ENSG00000259883.1	6108.999552
ENSG00000251883.3	0
ENSG00000260475.2	0
ENSG0000021788.1	0
ENSG00000167578.15	52440.130043
ENSG00000263089.1	34384.027373
ENSG00000172137.17	44565.5907
ENSG00000167700.7	22817.4208
ENSG00000167578.15	255959.2513
ENSG00000240423.1	130.5246249
ENSG00000167578.15	52440.130043
ENSG00000271616.1	0
ENSG00000167578.15	52440.130043
ENSG00000230403.1	394.475669
ENSG0000021105.1	15.8132762
ENSG00000167578.15	52440.130043
ENSG00000182741.8	453338.6046
ENSG00000269446.1	119.0847554

Games	P	Expression
ENIGMA0000024268.2	1.666,464.179	
ENIGMA0000027011.3	460,234.343	
ENIGMA0000027519.15	53446,30062	
ENIGMA0000027804.1	0	
ENIGMA0000027827.5	68,615,466	
ENIGMA0000028038.30	20,559,931,251	
ENIGMA0000028042.1	0	
ENIGMA0000028486.12	104,947,738	
ENIGMA0000028242.12	49,865,618	
ENIGMA0000025983.1	6,198,999,952	
ENIGMA0000028119.3	0	
ENIGMA0000028475.2	0	
ENIGMA0000027888.1	0	
ENIGMA0000028039.1	52,273,206	
ENIGMA0000028308.1	3,484,272,035.6	
ENIGMA0000027217.17	41,485,967	
ENIGMA0000027607.7	2,267,714,608	
ENIGMA0000027607.7	2,267,714,608	
ENIGMA0000028042.1	310,526,749	
ENIGMA0000026042.9	15,088,921.6	
ENIGMA0000027166.1	0	
ENIGMA0000028481.1	0	
ENIGMA0000028044.1	394,475,566	
ENIGMA0000028116.1	83,151,125.82	
ENIGMA0000028481.1	0	
ENIGMA0000028141.8	453,348,638	
ENIGMA0000028426.4	119,084,705.4	

Gene	Δ	Expression
ENSG00000234268.2	1658.464179	
ENSG00000270123.2	402.243443	
ENSG00000270615.5	52446.1006	
ENSG00000273842.1	0	
ENSG0000027435.5	68405.45626	
ENSG00000248033.0	25599.2351	
ENSG00000277575.4	0	
ENSG00000218486.12	104.947378	
ENSG00000219422.2	49865.56	
ENSG00000259831.1	6.98.100922	
ENSG00000219831.1	0	
ENSG00000240975.2	0	
ENSG00000217888.1	0	
ENSG00000210818.1	157330.256	
ENSG00000230009.1	3484.02773	
ENSG00000272317.2	44385.957	
ENSG00000267700.7	22673.44050	
ENSG00000267700.7	22673.44050	
ENSG00000240423.1	330.526749	
ENSG00000260491.9	15583.1216	
ENSG00000271636.1	0	
ENSG00000214881.1	0	
ENSG00000216000.1	394.47569	
ENSG00000212105.1	1383.31282	
ENSG00000214844.1	0	
ENSG00000238215.8	5538.4068	
ENSG000002694416.2	129.0840754	

Merged Sample Expression Data

Genes

SAMPLES

	0	1	2	3	4	5	6	7	8	9	...	60474	60475	60476	60477	60478	60479	60480	60481	60482	submitter_id
0	574548	2263.14	983212	69718	54834.9	19718.1	175853	735123	38662.4	233190	...	0	0	0	0	0	0	0	0	0	TCGA-04-1331-01A-01R-1569-13
1	352295	4592.37	663107	39745.4	36553.5	41147.1	241313	396423	37567	128693	...	0	0	0	0	0	0	0	0	0	TCGA-04-1332-01A-01R-1564-13
2	295162	649.026	1.21115e+06	57385.5	33097.4	58051.8	228615	346066	105567	408267	...	0	0	0	0	0	0	0	0	0	TCGA-04-1338-01A-01R-1564-13
3	329580	1835.59	1.08437e+06	33812.3	24516.1	22330.6	42134.4	895558	56178	83847.3	...	0	0	0	0	0	0	0	0	0	TCGA-04-1341-01A-01R-1564-13
4	289269	40061.7	2.44837e+06	26399.5	18248	49610	74761.1	571992	71951.9	98726.4	...	0	0	0	0	0	0	0	0	0	TCGA-04-1343-01A-01R-1564-13
...
4495	1.18093e+06	0	1.01139e+06	67877.2	15005.7	50527.3	6.21536e+06	1.47373e+06	459656	167488	...	0	0	0	0	0	0	0	0	0	TCGA-ZS-A9CD-01A-11R-A37K-07
4496	929228	0	869800	95607.5	17188.6	9352.12	7.61121e+06	196838	354465	138074	...	0	0	0	0	0	0	0	0	0	TCGA-ZS-A9CE-01A-11R-A37K-07
4497	469276	476.683	516938	110051	34469.4	37334.7	5.95811e+06	427832	323833	154861	...	0	0	0	0	0	0	0	0	0	TCGA-ZS-A9CF-01A-11R-A38B-07
4498	2.44119e+06	18282.7	853547	79288.7	106926	42593.9	4.80111e+06	955338	331924	177020	...	0	0	0	0	0	0	0	0	0	TCGA-ZS-A9CG-01A-11R-A37K-07
4499	259853	505.488	591328	74253.7	42553.5	118772	148978	508465	153862	170412	...	0	0	0	0	0	0	0	0	0	TCGA-ZX-AA5X-01A-11R-A42T-07

4500 rows × 60484 columns

Transpose and
add as a row

Genes	Expression
ENSG0000024298.2	3038.404179
ENSG0000027611.3	403.734143
ENSG0000026978.15	52440.1006
ENSG0000027840.1	0
ENSG0000028121.1	6885.4526
ENSG0000024293.10	25999.2351
ENSG0000025271.4	0
ENSG0000025446.12	154.947378
ENSG0000021842.12	406856.458
ENSG0000021881.1	6518.19052
ENSG0000021881.3	0
ENSG0000028471.2	0
ENSG0000026788.1	0
ENSG0000023438.11	95730.2056
ENSG0000023438.1	2484.03713
ENSG0000027121.17	41485.9507
ENSG00000257780.7	22672.4208
ENSG0000023444.2	2982.24035
ENSG0000024042.1	335.5246749
ENSG0000026041.9	125863.5216
ENSG00000271816.1	0
ENSG00000214881.1	0
ENSG00000238041.1	384.475669
ENSG00000231101.1	1583.112582
ENSG0000024044.1	0
ENSG00000252141.8	45338.40648
ENSG00000289416.4	119.0847054
ENSG00000254911.1	0

Quantifying mRNA abundance and Scaling

- GDC harmonization data is provided in FPKM-UQ
- In our code, FPKM-UQ is rescaled to TPM using the following formula.

$$\text{TPM}_i = \left(\frac{\text{FPKM}_i}{\sum_j \text{FPKM}_j} \right) \cdot 10^6$$

- TPM has nice mathematical properties and a stable entity

<https://docs.gdc.cancer.gov/Encyclopedia/pages/HTSeq-FPKM-UQ/>

Mapping and quantifying mammalian transcriptomes
by RNA-Seq

Ali Mortazavi^{1,2}, Brian A Williams^{1,2}, Kenneth McCue¹, Lorian Schaeffer¹ & Barbara Wold¹

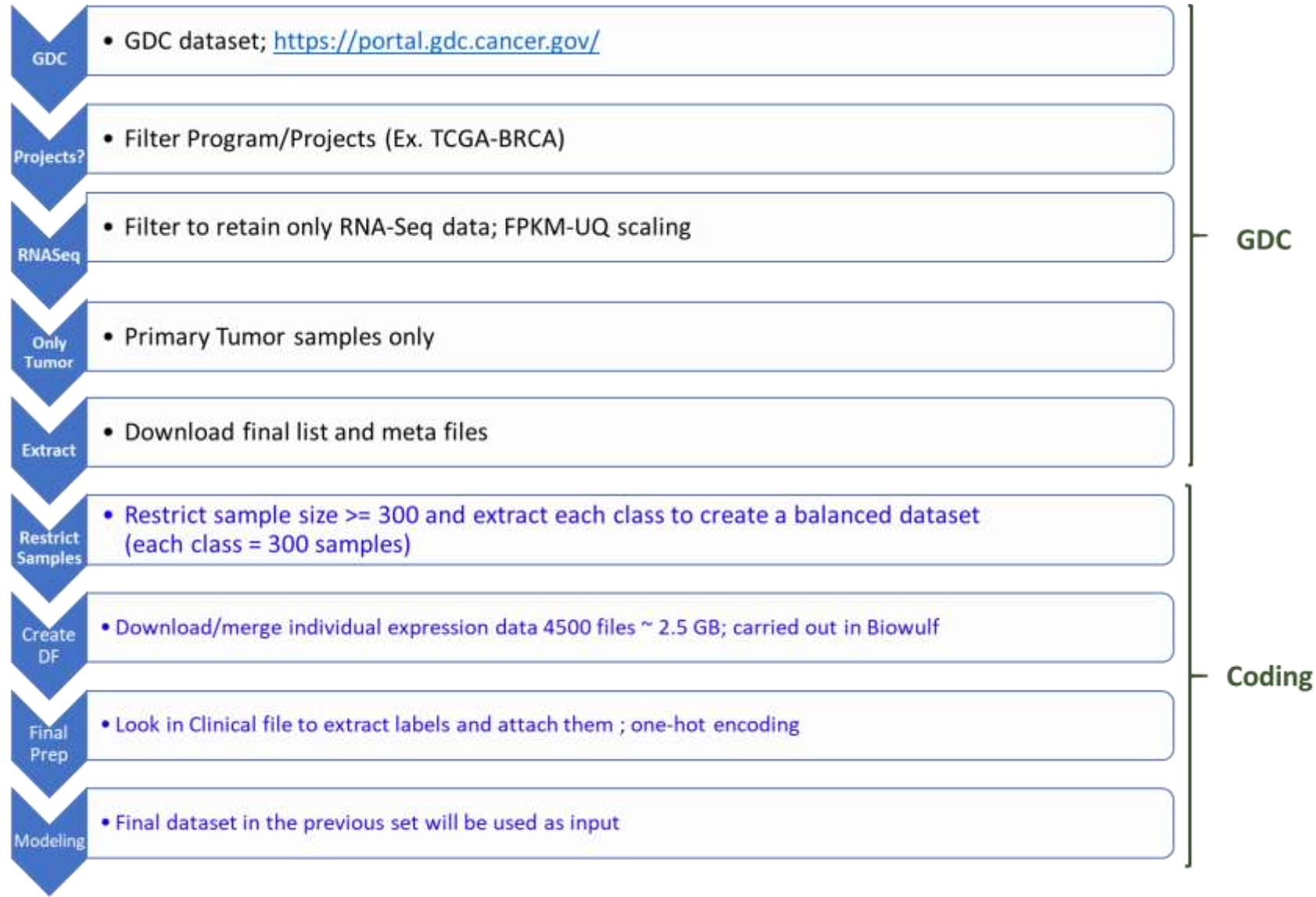
One-hot encoding to convert Cancer types to numbers

- Convenient to transform categorical variables into a numerical quantity for computations
 - BRCA to 0 ; LUAD to 1 etc.
 - 0, 1, 2, 3, ..., 13, 14, 14

TCGA-CESC
TCGA-LIHC
TCGA-STAD
TCGA-OV
TCGA-BLCA
TCGA-THCA
TCGA-PRAD
TCGA-COAD
TCGA-KIRC
TCGA-LUSC
TCGA-HNSC
TCGA-LGG
TCGA-LUAD
TCGA-UCEC
TCGA-BRCA

```
>>> encoded
array([[1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1.]],
      dtype=float32)
```


Data preparation steps summary



Before we break for hands-on

- **Python as the programming language for this workshop, but similar libraries are available in R or other languages**



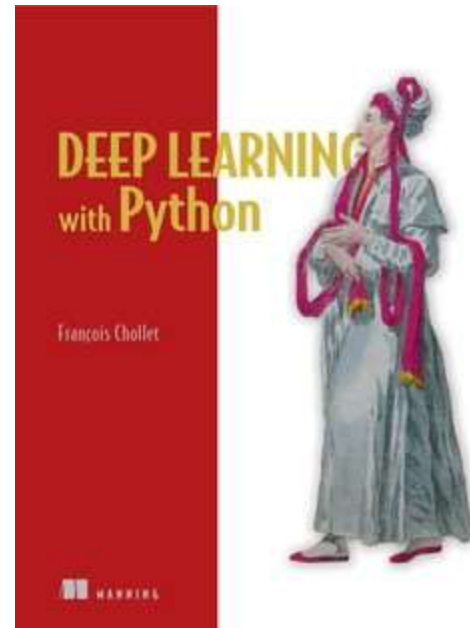
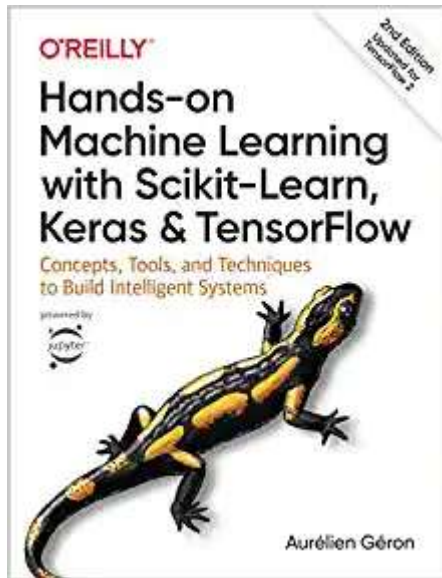
- **Will use Jupyter Notebook for sharing the code**
 - With little effort one can convert the Python code into R and still use Jupyter Notebook

To be continued after hands-on

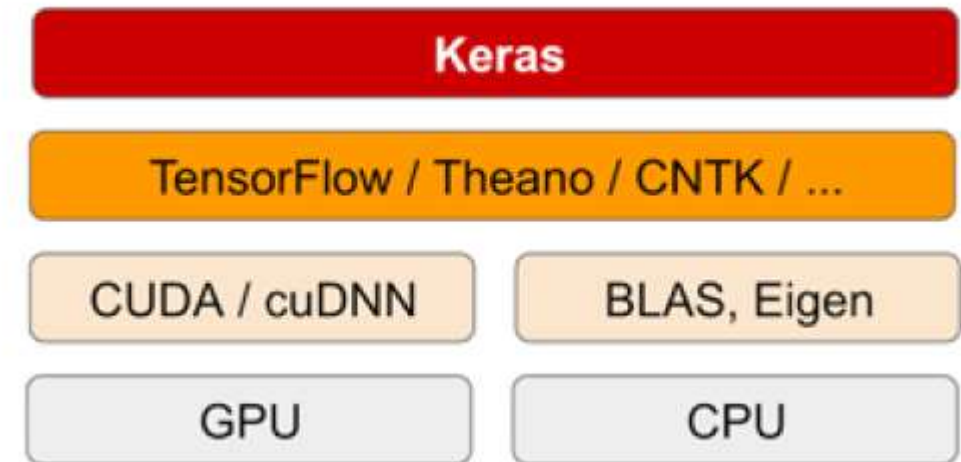
<https://github.com/ravichas/ML-TC1>

Before we begin the modeling section ...

- Due to lack of time, I won't be covering the basics of Neural Network



Keras is a high-level NN package that is built on top of popular high-level libraries (TF, Theano). Works well with CPU/GPU



These are good books for beginners and up

Figure from Deep Learning with Python

Supervised Learning

- Goal
 - Construct a model that takes in input features/target pair to return a prediction for target/outcome
- Train a machine learning
 - Model refers to learning its parameters, which typically involves minimizing a loss function on training data with the aim of making accurate predictions on unseen (test) data

Supervised Learning:

Data: (x,y) ; where x is the genomic expression profile ; y is the cancer classes

Goal? Learn the function that maps
 $x \rightarrow y$

Terminology

	0	1	2	3	4	5	6	7	8	9	...	60474	60475	60476	60477	60478	60479	60480	60481	60482	submitter_id
0	574548	2263.14	983212	69718	54834.9	19718.1	175853	735123	38662.4	233190	...	0	0	0	0	0	0	0	0	0	TCGA-04-1331-01A-01R-1569-13
1	352295	4592.37	663107	39745.4	36553.5	41147.1	241313	396423	37567	128693	...	0	0	0	0	0	0	0	0	0	TCGA-04-1332-01A-01R-1564-13
2	295162	649.026	1.21115e+06	57385.5	33097.4	58051.8	228615	346066	105567	408267	...	0	0	0	0	0	0	0	0	0	TCGA-04-1338-01A-01R-1564-13
3	329580	1835.59	1.08437e+06	33812.3	24516.1	22330.6	42134.4	895558	56178	83847.3	...	0	0	0	0	0	0	0	0	0	TCGA-04-1341-01A-01R-1564-13
4	289269	40061.7	2.44837e+06	26399.5	18248	49610	74761.1	571992	71951.9	98726.4	...	0	0	0	0	0	0	0	0	0	TCGA-04-1343-01A-01R-1564-13

- **Columns**
 - input variables or features or attributes
- **Outcome column**
 - Outcome variables or targets
- **Rows**
 - Training example or instance
- **Whole table Training data set**

What is different about Neural Network?

- If you know the equation (algorithm), then you feed in the **input** and you get the **output**.
You can code the function yourself

```
def function(x):  
    y = 2.0 + 5.0 * x  
    return(y)
```

- You can choose to use linear modeling and use the data to figure the relationship

```
Model ← lm( y ~ x)
```

- Neural Network using the data learn the algorithm.

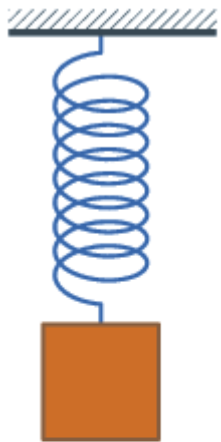
INPUT

ALGORITHM

OUTPUT

A Simple Network

Input: Mass or M (kg)
Output: Length or L (m)

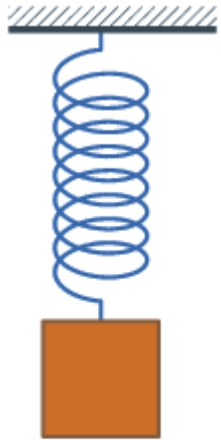


M	L
Input	Output
0.125	0.39
0.25	0.40
0.5	0.43
1	0.48
2	0.58
3	???

[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

Based on Mary Attenborough, in [Mathematics for Electrical Engineering and Computing](#), 2003

A Simple Network



M	L
0.125	0.39
0.25	0.40
0.5	0.43
1	0.48
2	0.58
3	0.68

$$L = 0.1 * Mass + 0.38$$

[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

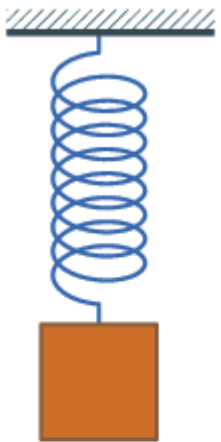
Mary Attenborough, in [Mathematics for Electrical Engineering and Computing](#), 2003

A Simple Network

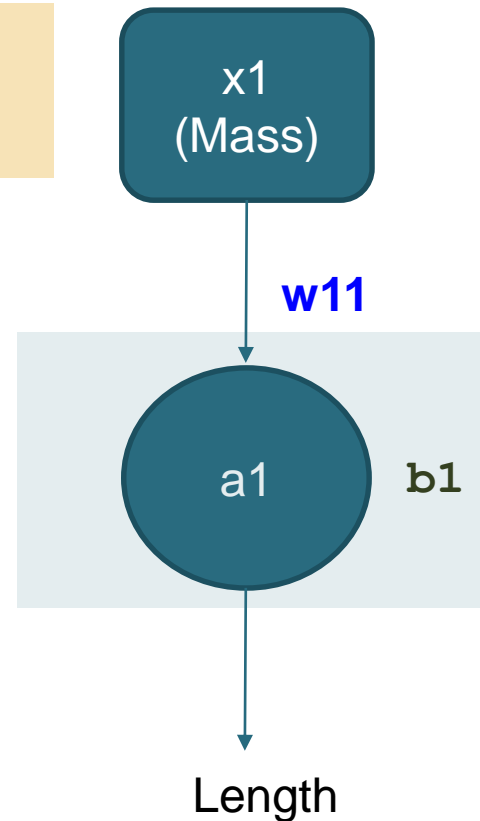
$$a1 = x1 * w11 + b1$$

$$L = M * 0.1 + 0.38$$

[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)



Hidden Layer



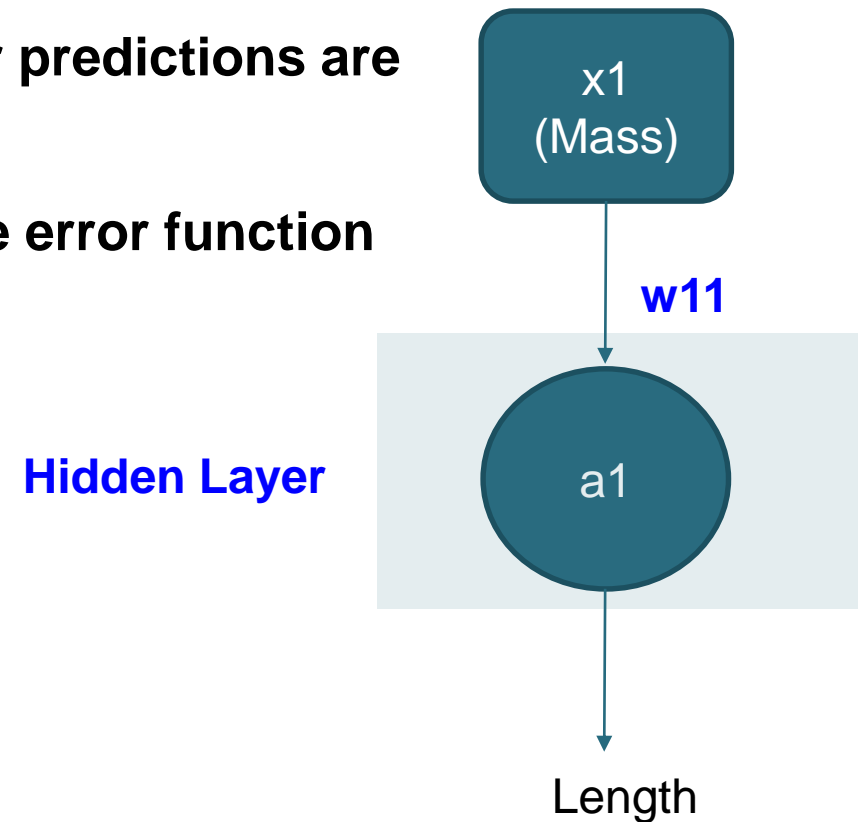
M	L
0.125	0.39
0.25	0.40
0.5	0.43
1	0.48
2	0.58
3	0.68

These are the model variables: `[array([[0.10058284]], dtype=float32), array([0.37793916], dtype=float32)]`

Based on Mary Attenborough, in [Mathematics for Electrical Engineering and Computing](#), 2003

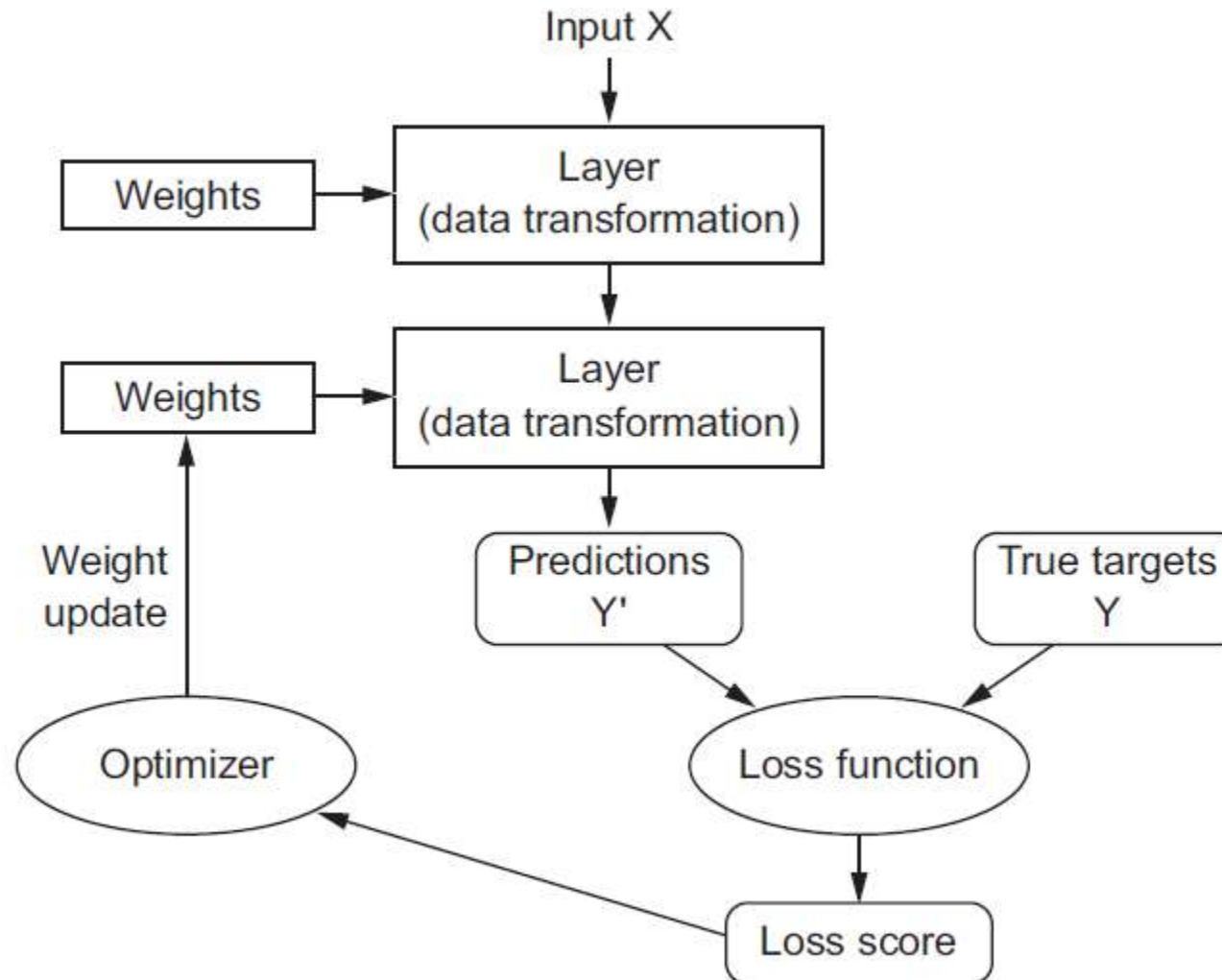
Error minimization

- Goal is to choose W s such that predictions of the network should be close to y
- Error function or cost function a measure how good our predictions are
- Eventually, we want to pick a set of w that minimizes the error function



Deep Learning Procedure

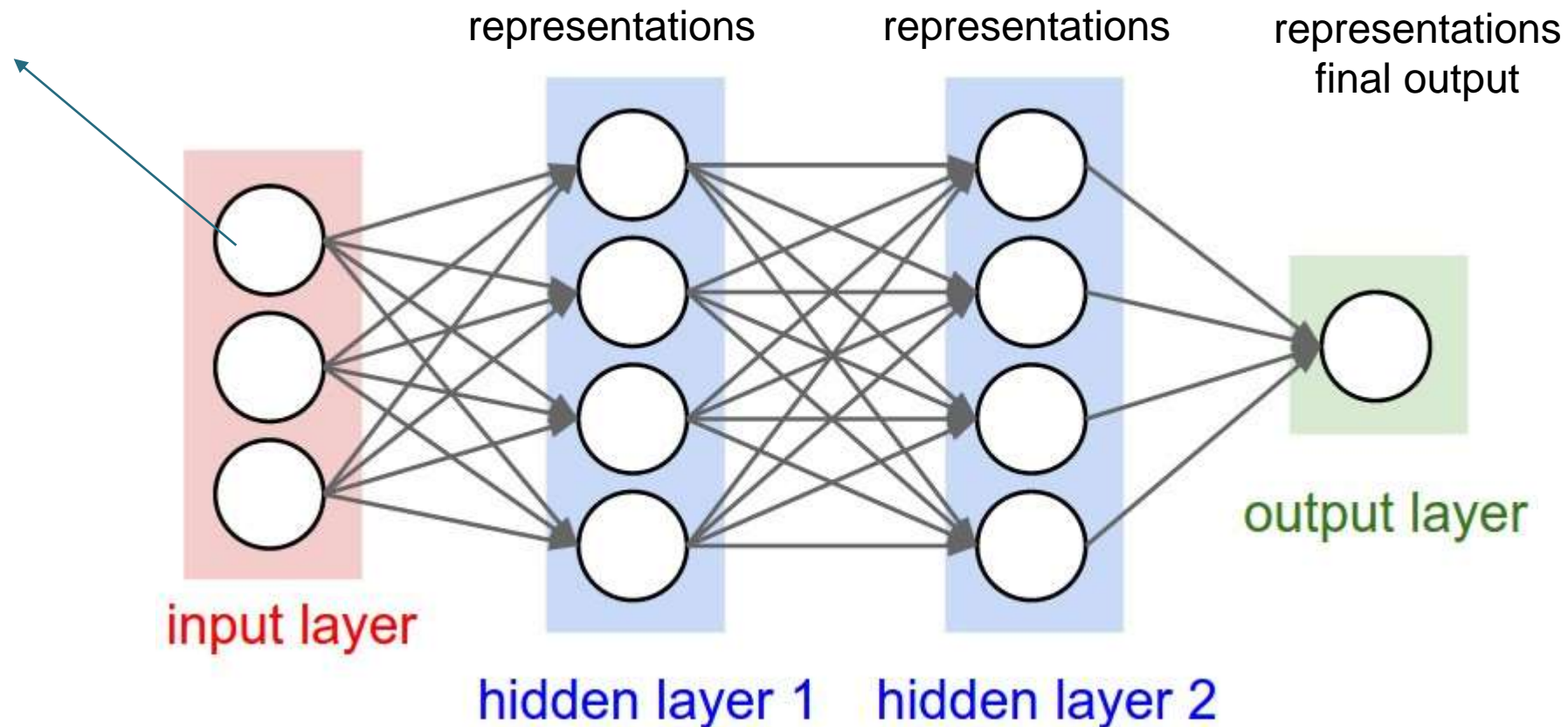
Taken from Deep Learning with Keras book



Vanilla network

Each neuron receives input from all the neurons in the previous layer (densely connected)

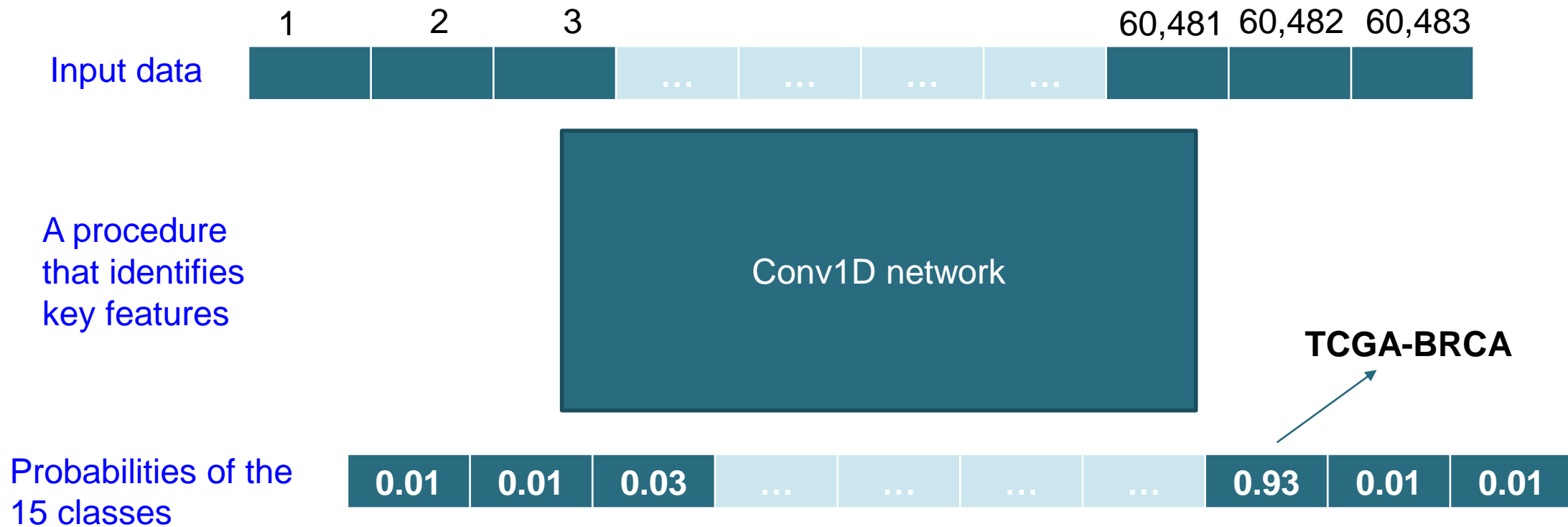
Neuron: a unit that holds a number



[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)

Convolutional Neural Network

- We are going to take a vector of genomic expression values and feed them into a network with a series of operations to create a model
- Model is what we call convolutional-1D network

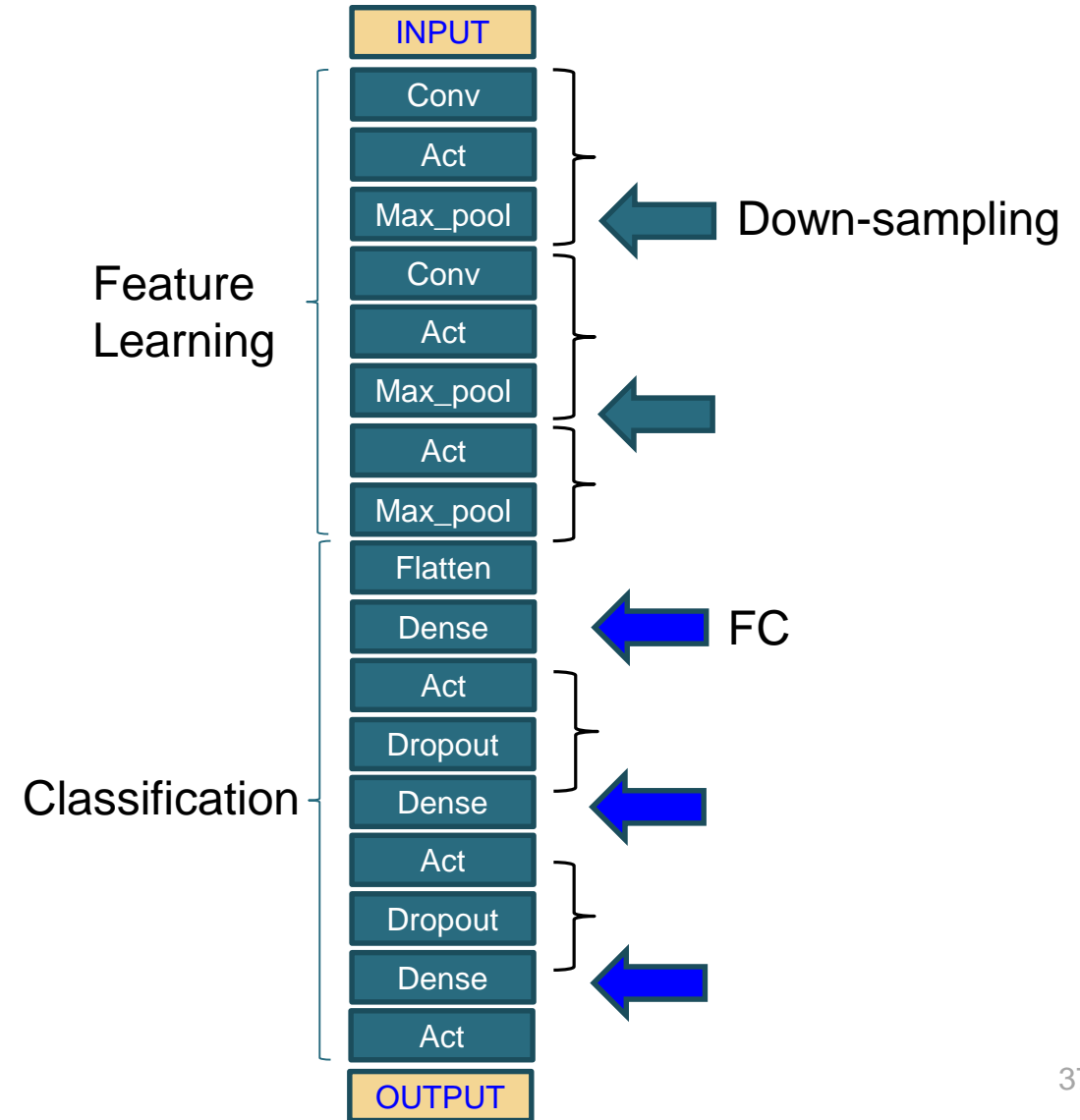


Components of conv1D

1. **Act: Activation**
2. **Conv: Convolution**
3. **Max_pool: Maxpooling**
4. **Flatten**
5. **Dense**
6. **Dropout**

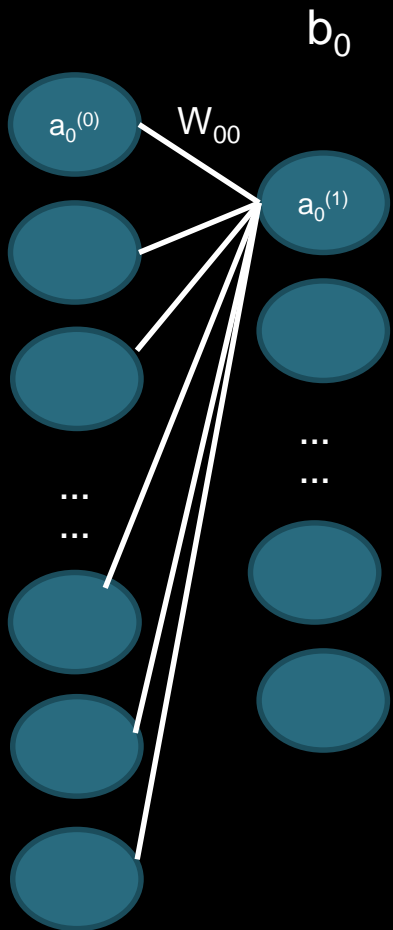
Topology of a network defines a “hypothesis space”

Choosing a specific topology is usually not straightforward and comes with practice.



1. Activation function

$$a^{(L)} = \sigma(w^{(L)}a^{(L-1)} + b^{(L)})$$



$$\sigma \left(\begin{bmatrix} W_{0,0} & W_{0,1} & \dots & W_{0,n} \\ W_{1,0} & W_{1,1} & \dots & W_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{k,0} & W_{k,1} & \dots & W_{k,n} \end{bmatrix} \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \right)$$

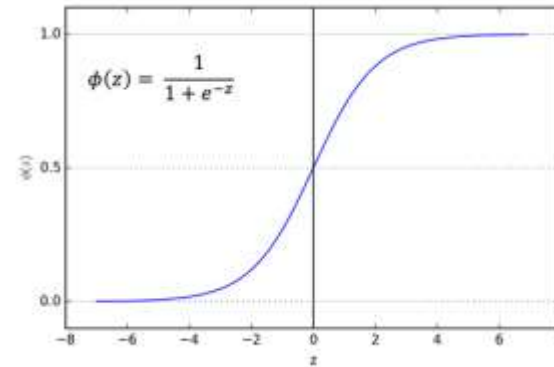
$$a_0^{(1)} = \sigma(W_{00}a_0^{(0)} + W_{0,1}a_1^{(0)} + \dots W_{0,n}a_n^{(0)} - b_0)$$

1. Activation Function

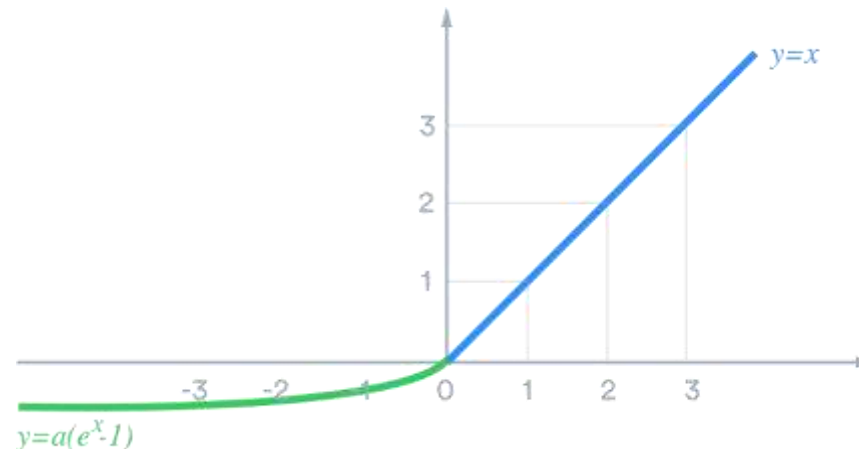
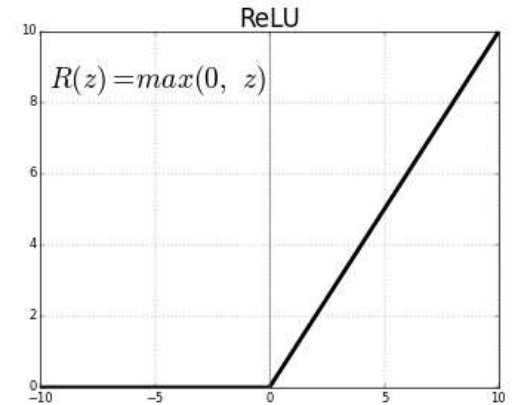
- Activation functions are included to create non-linearity

[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

- Sigmoid
- ReLU
- Leaky ReLU
- ELU
- Maxout
- Tanh



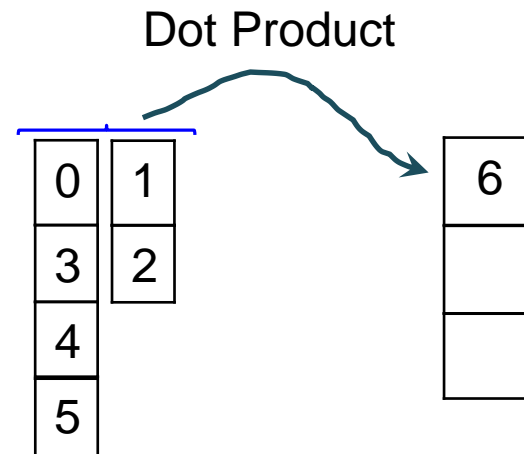
Squashes the #s to [0, 1]



2. Convolution

Process of applying filter (kernel) to the data for the purpose of subsampling. Kernel is a matrix that has a smaller dimension than the input data creates chunks

Reduces the number of parameters and allow creation of deeper networks

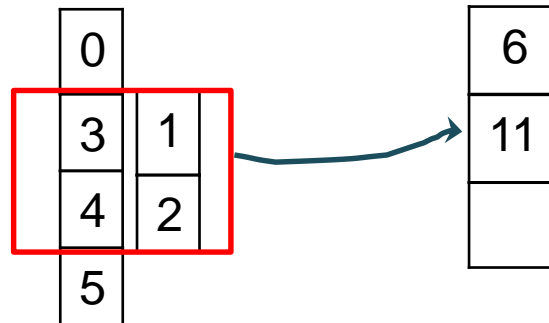


2. Convolution

Process of applying filter (kernel) to the data for the purpose of subsampling. Kernel is a matrix that has a smaller dimension than the input data creates chunks

Reduces the number of parameters and allow creation of deeper networks

Dot Product

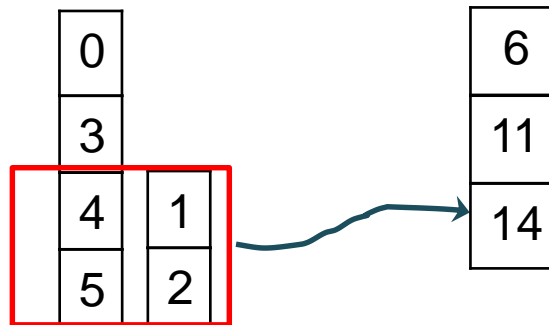


2. Convolution

Process of applying filter (kernel) to the data for the purpose of subsampling. Kernel is a matrix that has a smaller dimension than the input data creates chunks

Reduces the number of parameters and allow creation of deeper networks

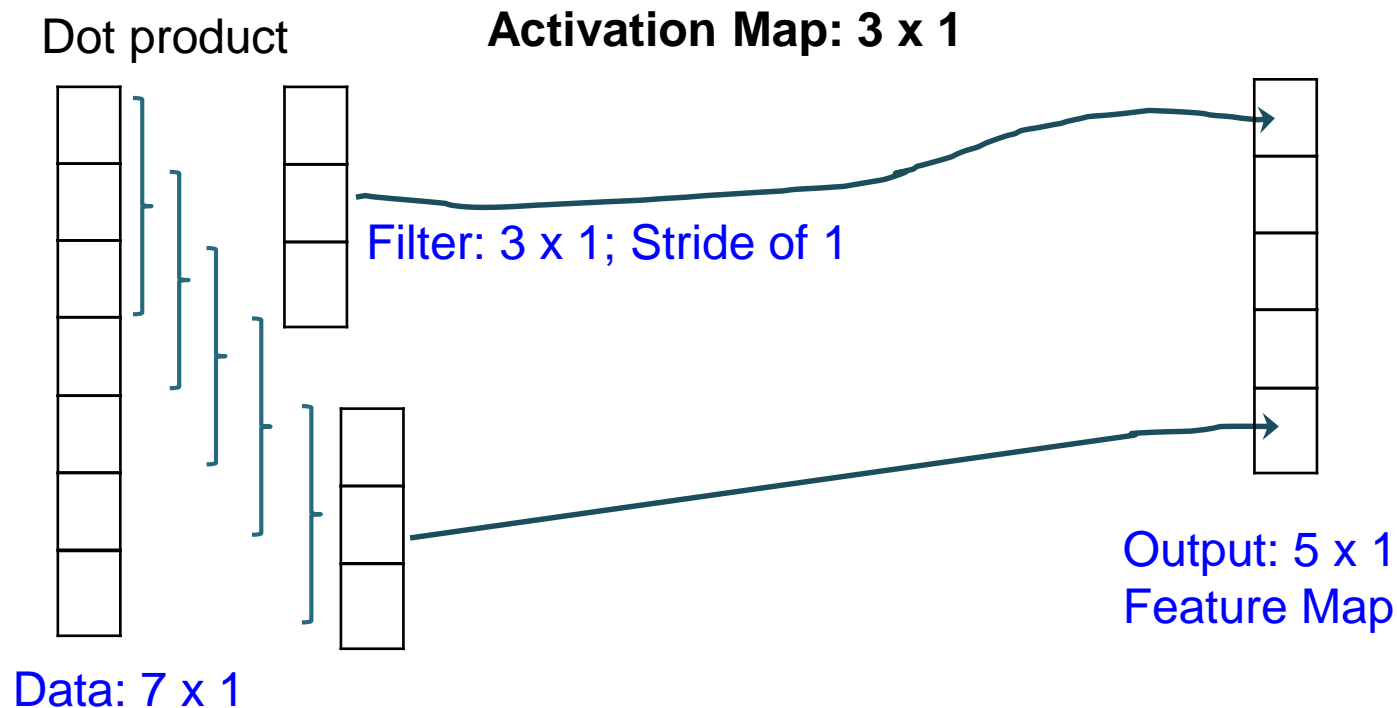
Dot Product



2. Convolution

Process of applying filter (kernel) to the data for the purpose of subsampling. Kernel is a matrix that has a smaller dimension than the input data creates chunks

Reduces the number of parameters and allow creation of deeper networks



$(N-F)/\text{stride}+1$ will be the size after filtering

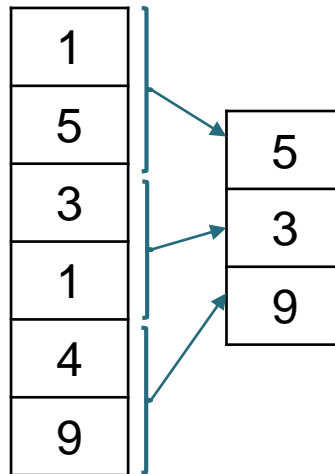
$(7-3)/1+1 = 5$;
zero padding on the border

2. Convolution

- **Convolution Layer**
 - Hyperparameters
 - Number of filters
 - Spatial extent
 - Stride
 - Amount of zero padding

3. Pooling

- Pooling makes the representations smaller/manageable (downsampling) by retaining only important features; creates smaller clusters of manageable size
- Each activation map will be pooled separately.
- Common approach is Max Pooling



Max-pooling
with filter size
of 2x1 and
stride of 2

Max Pooling Intuition:

Enhancing the signals by looking at a region and pick the maximum activation value

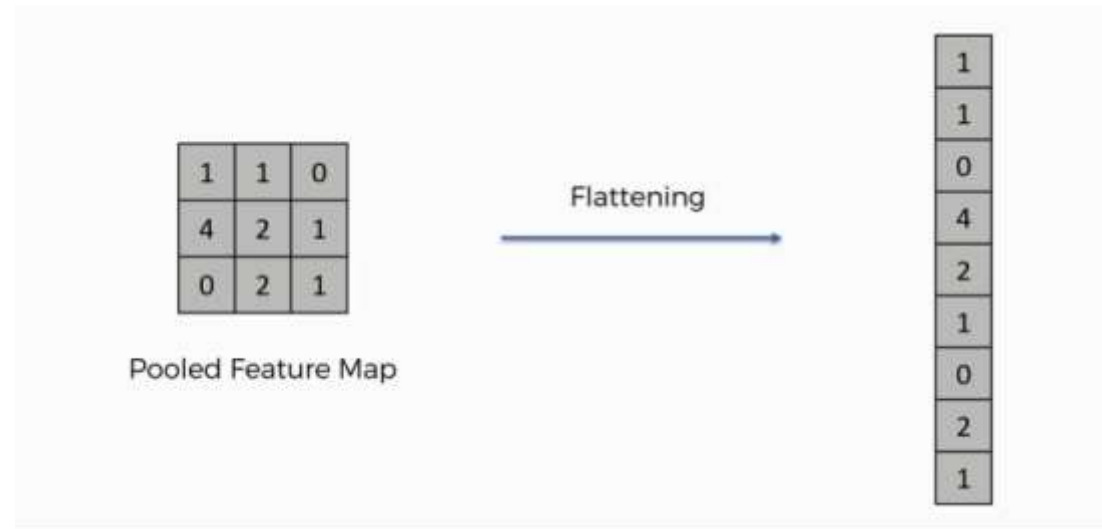
Each of these are activation and we are looking for

Research shows that zero-padding is not followed.
Because we are interested in down-sampling

Common setting for filter 2 or 3

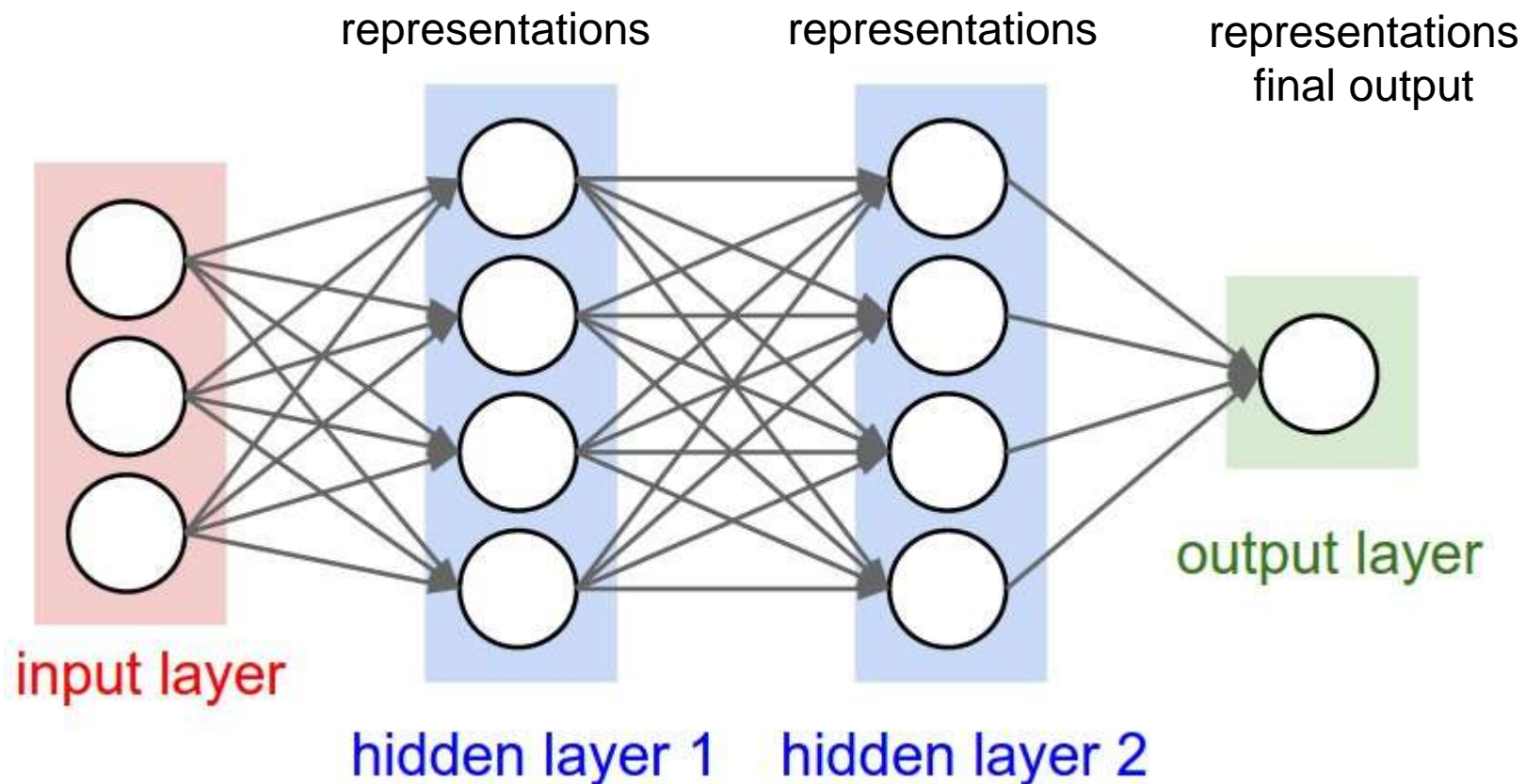
4. Flatten

Procedure to transform a 2D matrix (features) to a 1D vector which in turn can be fed into a fully-connected layer (dense)



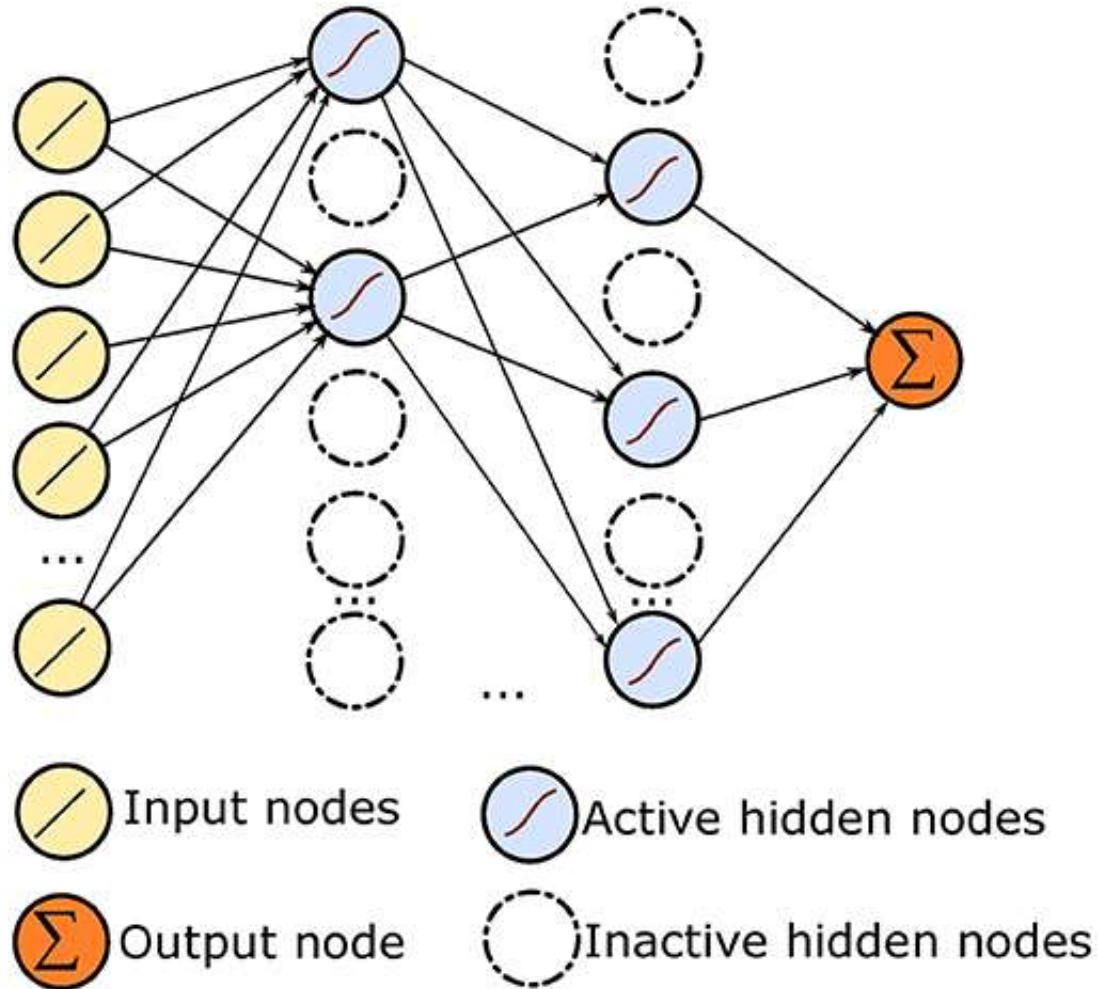
5. Dense

Each neuron receives input from all the neurons in the previous layer (densely connected)



[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)

6. Dropout



Imbalance in the weights among the nodes can lead to some node weights not contributing to the learning

**One solution:
Remove a random proportion of selection of neurons in a neural network during training**

Can help weak learners become strong learners

6. Dropout



[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)

Model Summary

~ 154 M parameters

```
1.0 128 10 1
Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 60464, 128)	2688
activation_1 (Activation)	(None, 60464, 128)	0
max_pooling1d_1 (MaxPooling1D)	(None, 60464, 128)	0
conv1d_2 (Conv1D)	(None, 60455, 128)	163968
activation_2 (Activation)	(None, 60455, 128)	0
max_pooling1d_2 (MaxPooling1D)	(None, 6045, 128)	0
flatten_1 (Flatten)	(None, 773760)	0
dense_1 (Dense)	(None, 200)	154752200
activation_3 (Activation)	(None, 200)	0
dropout_1 (Dropout)	(None, 200)	0
dense_2 (Dense)	(None, 20)	4020
activation_4 (Activation)	(None, 20)	0
dropout_2 (Dropout)	(None, 20)	0
dense_3 (Dense)	(None, 15)	315
activation_5 (Activation)	(None, 15)	0
Total params: 154,923,191		
Trainable params: 154,923,191		
Non-trainable params: 0		



Code execution and progress

```
Epoch 00001: val_loss improved from inf to 2.56791, saving model to Pilot1.h5

Epoch 2/400
3375/3375 [=====] - 228s 68ms/step - loss: 2.2202 - acc: 0.2821 - val_loss: 1.8444 - val_acc:
Epoch 00002: val_loss improved from 2.56791 to 1.84441, saving model to Pmodel.h5

Epoch 3/400
3375/3375 [=====] - 228s 68ms/step - loss: 1.4736 - accuracy: 0.5206 - val_loss: 0.9554 - val_acc:
Epoch 00003: val_loss improved from 1.84441 to 0.95540, saving model to Pmodel.h5

Epoch 4/400
3375/3375 [=====] - 228s 68ms/step - loss: 0.8795 - accuracy: 0.7058 - val_loss: 0.4835 - val_acc:
Epoch 00004: val_loss improved from 0.95540 to 0.48347, saving model to Pmodel.h5

Epoch 5/400
3375/3375 [=====] - 228s 68ms/step - loss: 0.5968 - accuracy: 0.8107 - val_loss: 0.4083 - val_acc:
Epoch 00005: val_loss improved from 0.48347 to 0.40829, saving model to Pmodel.h5

Epoch 6/400
3375/3375 [=====] - 228s 68ms/step - loss: 0.4529 - accuracy: 0.8519 - val_loss: 0.3236 - val_acc:
Epoch 00006: val_loss improved from 0.40829 to 0.32363, saving model to Pmodel.h5

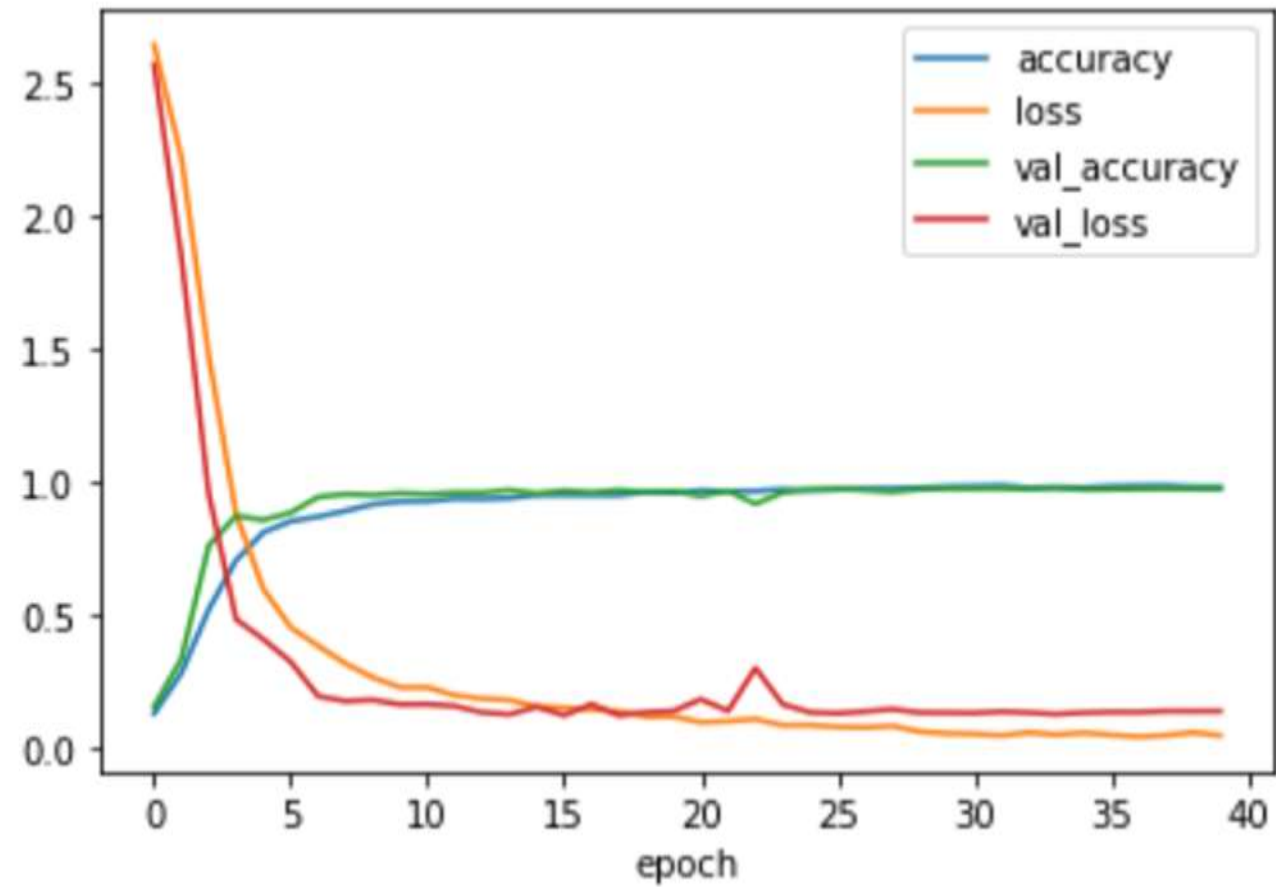
Epoch 7/400
3375/3375 [=====] - 228s 68ms/step - loss: 0.3835 - accuracy: 0.8690 - val_loss: 0.1944 - val_acc:
Epoch 00007: val_loss improved from 0.32363 to 0.19439, saving model to Pmodel.h5

Epoch 8/400
3375/3375 [=====] - 228s 68ms/step - loss: 0.3170 - accuracy: 0.8910 - val_loss: 0.1754 - val_acc:
Epoch 00008: val_loss improved from 0.19439 to 0.17536, saving model to Pmodel.h5

Epoch 9/400
3375/3375 [=====] - 228s 67ms/step - loss: 0.2647 - accuracy: 0.9156 - val_loss: 0.1800 - val_acc:
Epoch 00009: val_loss did not improve from 0.17536

Epoch 10/400
3375/3375 [=====] - 228s 68ms/step - loss: 0.2276 - accuracy: 0.9265 - val_loss: 0.1632 - val_acc:
Epoch 00010: val_loss improved from 0.17536 to 0.16323, saving model to Pmodel.h5
```


Model Performance



Thank you!

[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Questions/Comments

S. Ravichandran
ravichandrans@mail.nih.gov

