Predicting diseases through machine learning models

John D. Lewis, Ph.D.

Bird Dogs Chair in Translational Oncology Research

Professor, Department of Oncology
University of Alberta

CanPath
Canadian Partnership for Tomorrow's Health

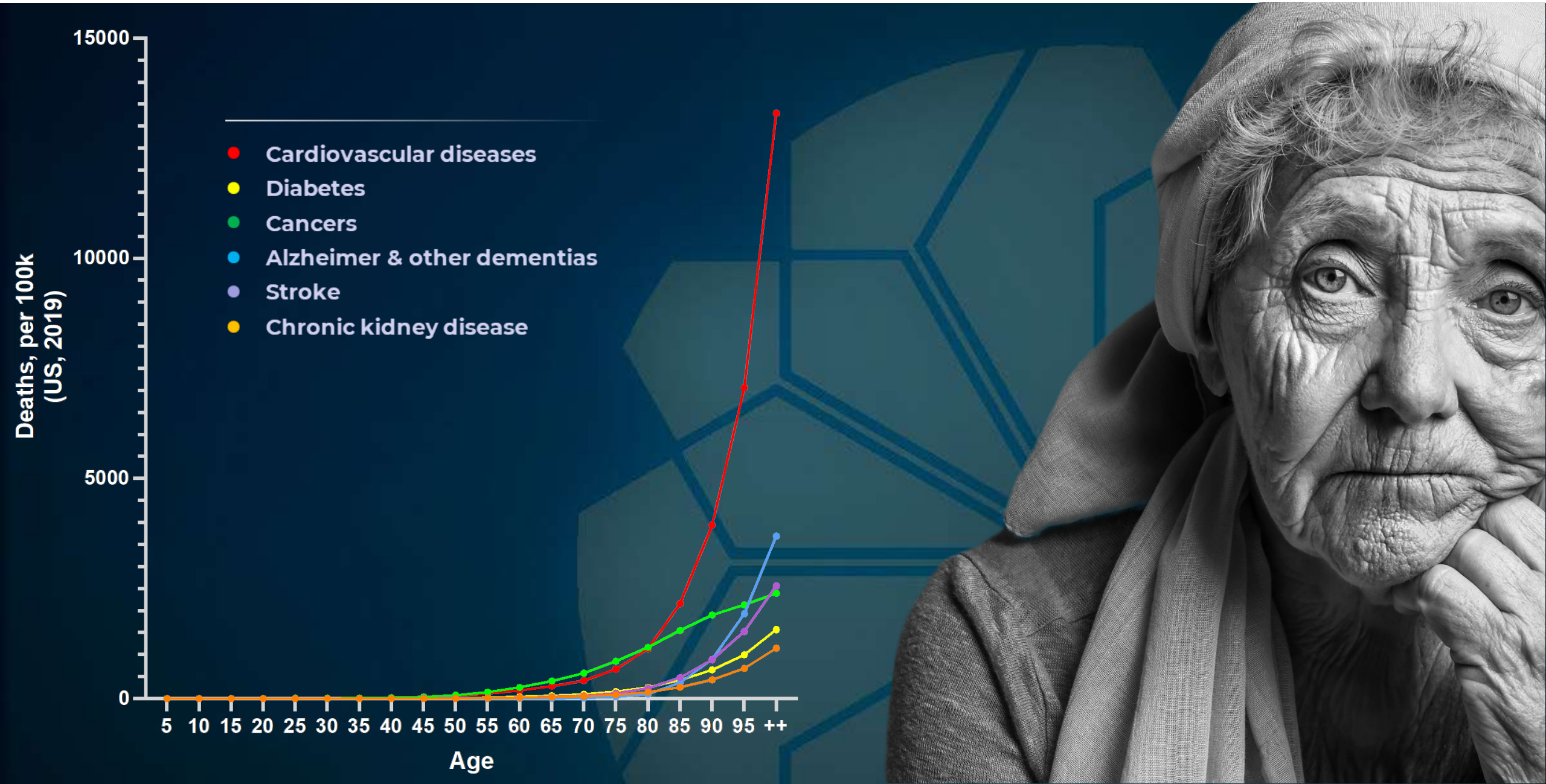
ALBERTA
PROSTATE CANCER
RESEARCH INITIATIVE
knowledge | action | impact

Disclosures

Founder and CEO of Nanostics Inc

Founder and CEO of Entos Pharmaceuticals
Founder and CSO of OncoSenX
Founder and CEO of Aegis Life
CSO at Oisin Biotechnologies

Incidence of chronic disease increases as we age...



… and carries a significant burden.

- As much as 50% of the global burden of disease is chronic illness (WHO).
- Chronic disease is a significant concern for Canadians, with one-half (51.6%) of the population over the age of 20 having one or more chronic diseases.
- Chronic diseases are estimated to cost Canadians $68 billion in direct healthcare costs and $122 billion in productivity losses.

Thanks to CanPath, we have access to BIG HEALTH DATA

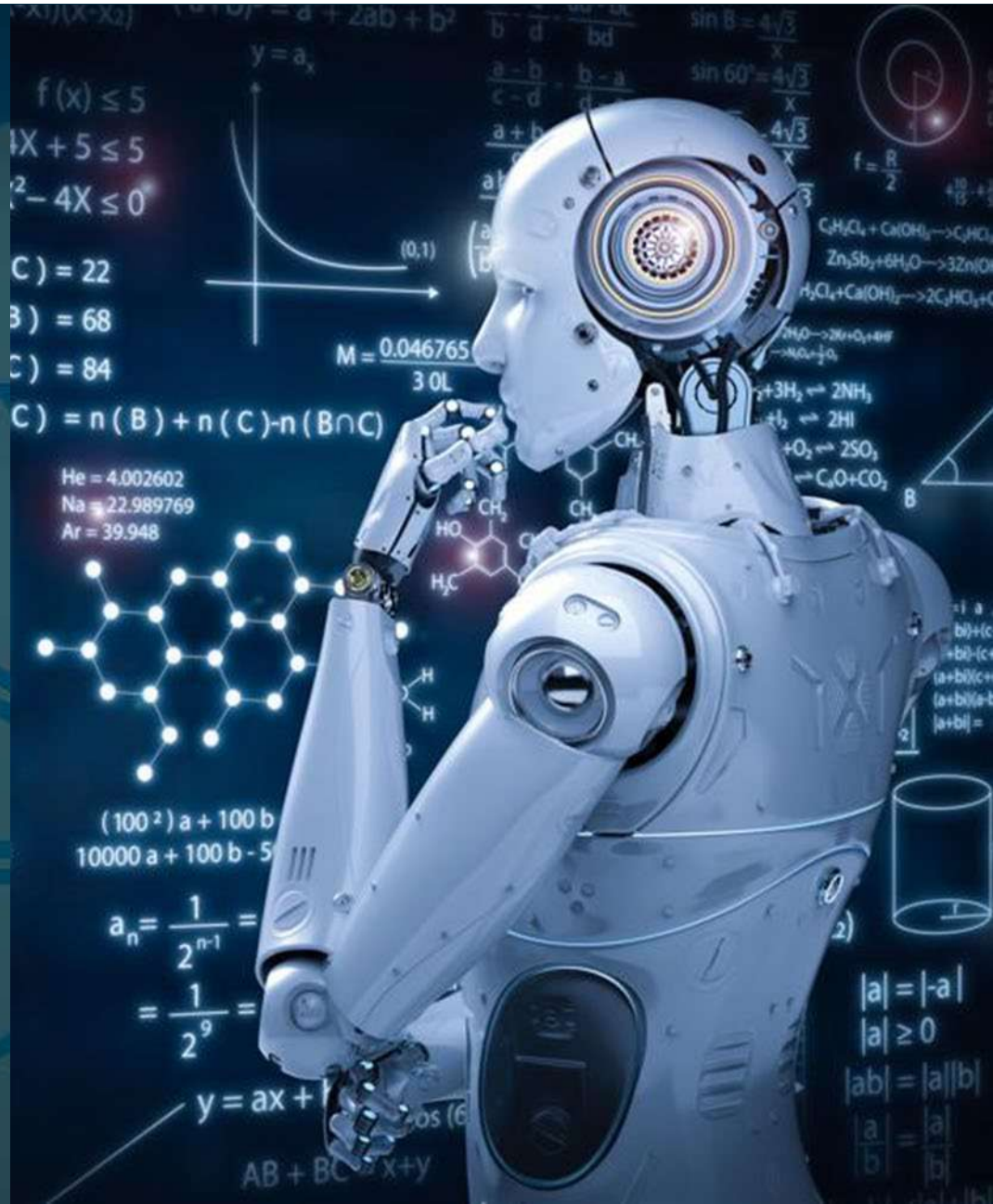
- Comprehensive genomic, clinical, behavioural and environmental data on 330,000 Canadians
- Baseline and follow-up data from five regional cohorts have been harmonized across the country, creating a pan-Canadian resource of more than 2,800 measures of participant health and lifestyle factors.
- Value is increasing over time as new data are added, technology advances, and incident health outcomes are captured.

How do we find patterns in thousands of measures over hundreds of thousands of individuals?

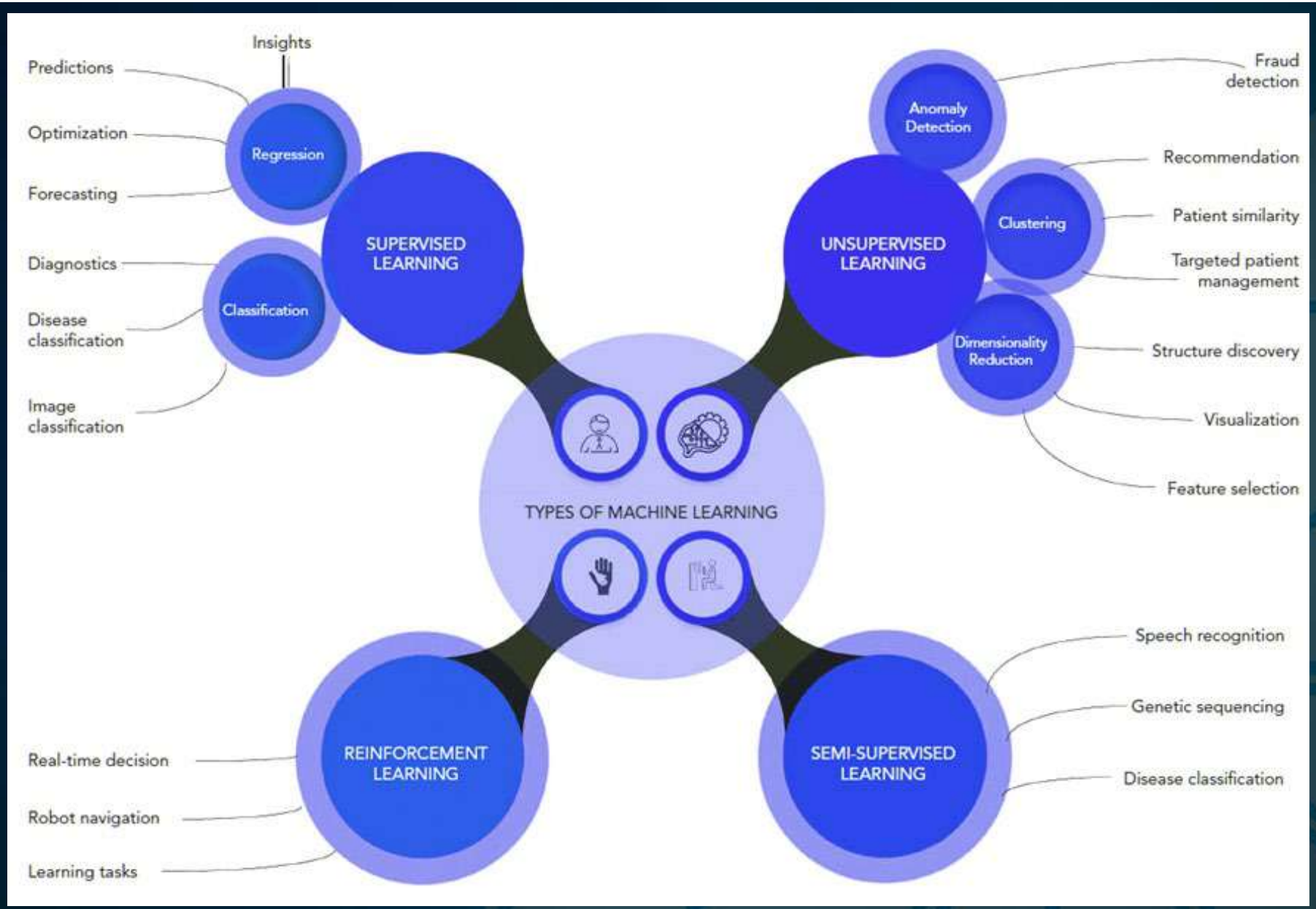
Machine learning (ML or AI)?

- The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.

– Oxford Languages, September 29, 2021



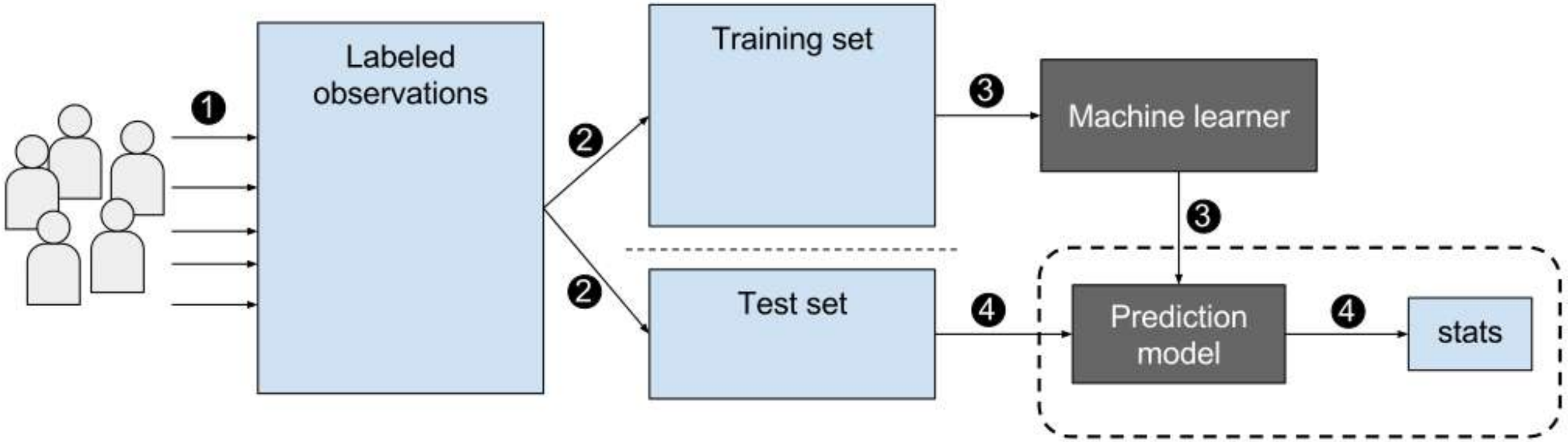
Types of machine learning



US Cardiology Review 2019;13(2):110–6

Machine learning (ML or AI)?

- Algorithms which use features (e.g., age or lab results) to predict labels (e.g., “no cancer” or “has cancer”) on observations (e.g., patients) which have known values for features and labels.



Machine learning algorithm can make predictions on <u>new</u> data

Many different machine learning algorithms

Linear models

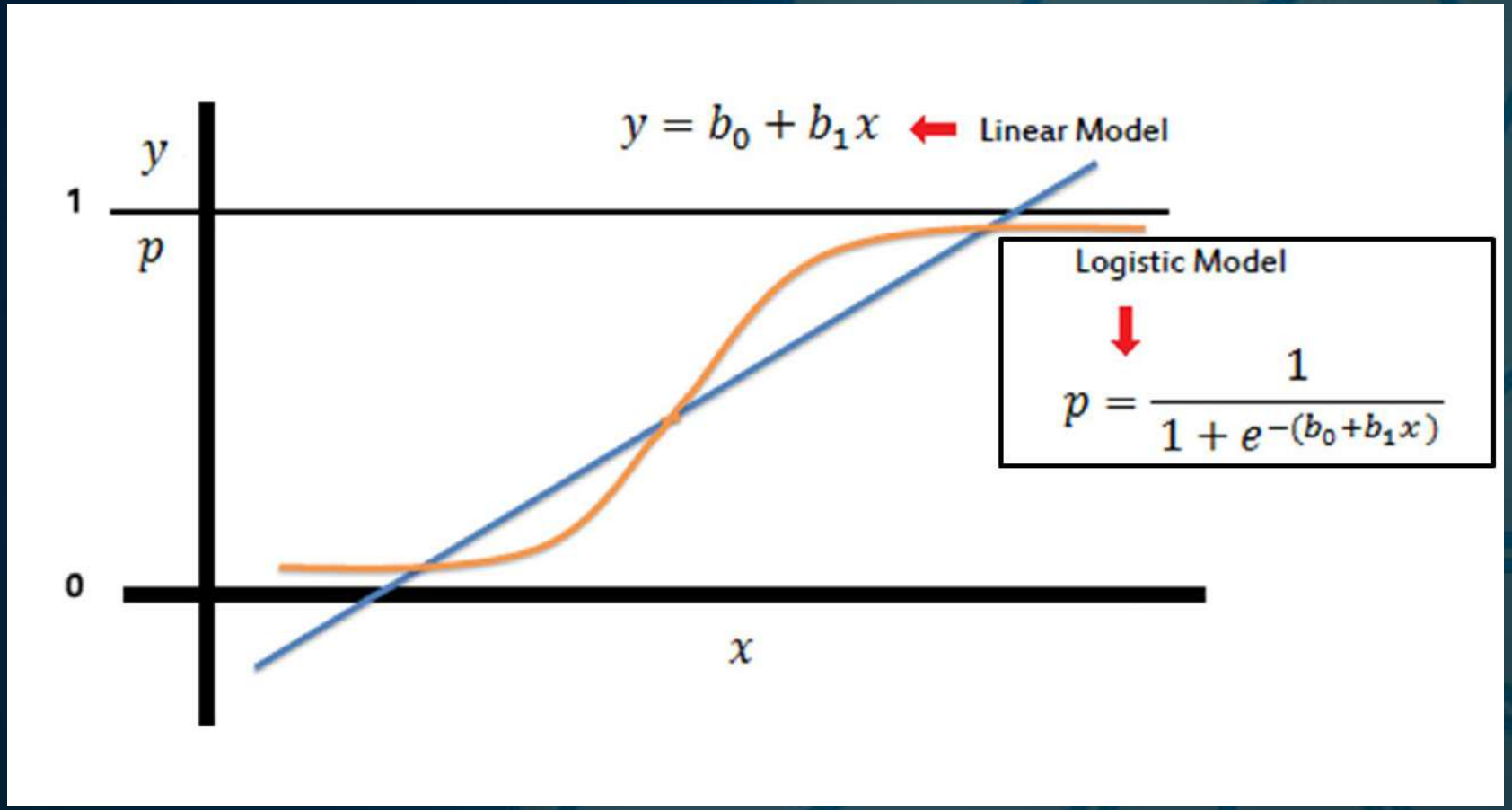
- Logistic regression
- Linear discriminant analysis
- Linear support vector machines

Non-linear models

- K-nearest neighbors
- Decision trees
- Neural networks

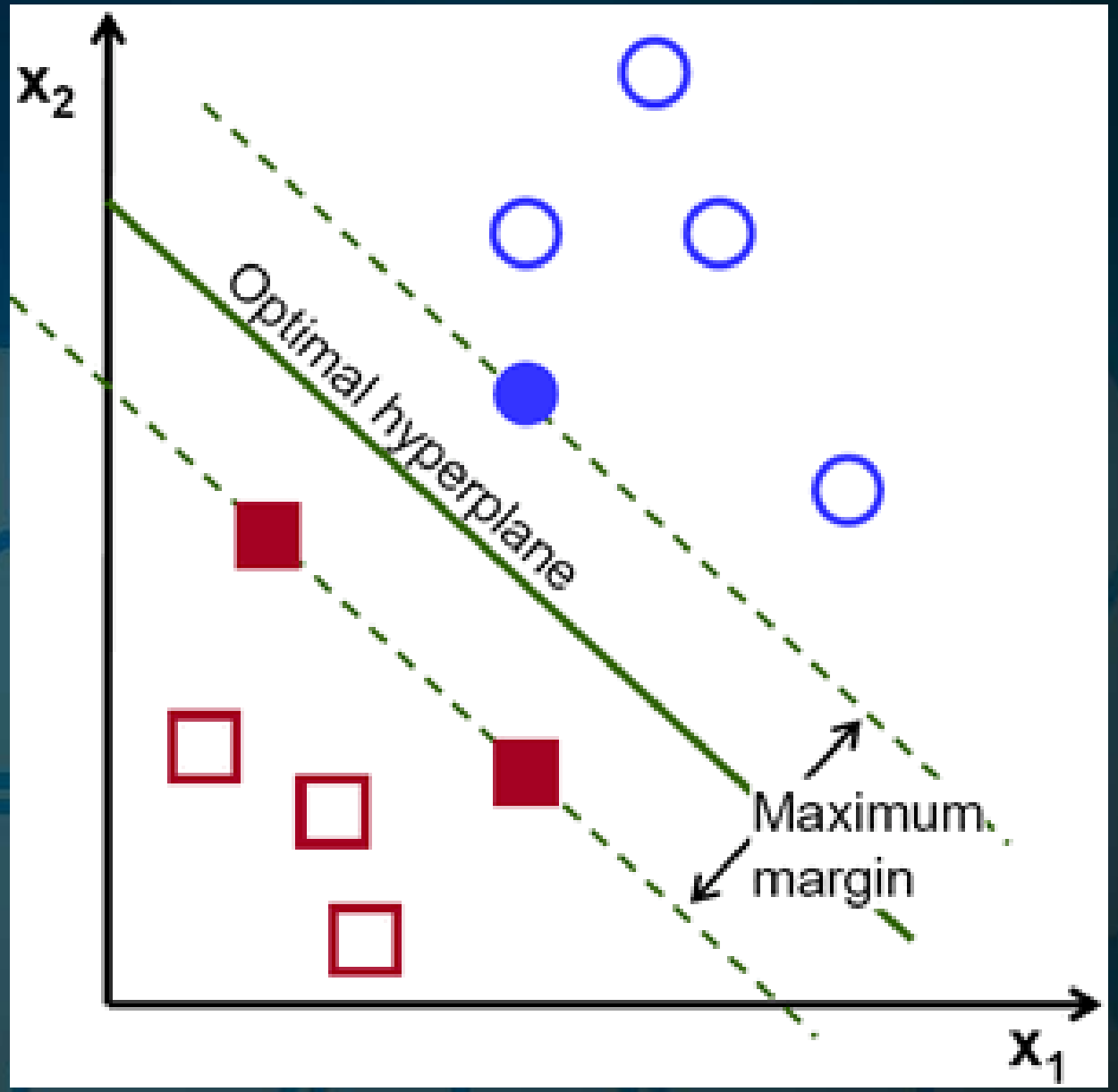
Logistic regression

- Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set.



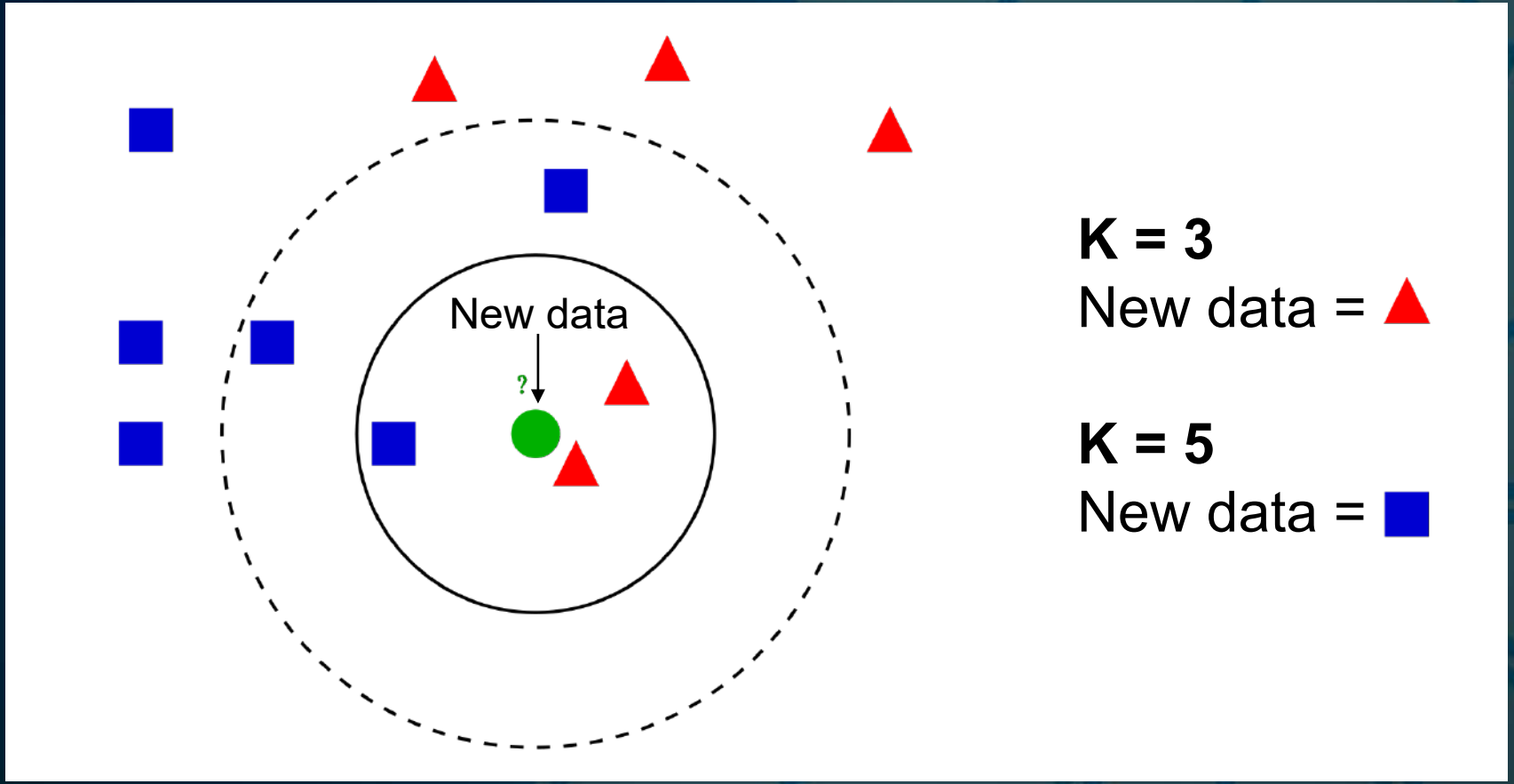
Support Vector Machines (SVM)

- Identifies the hyperplane with maximal distance from points in each group. Points on each side of hyperplane belong to separate groups.



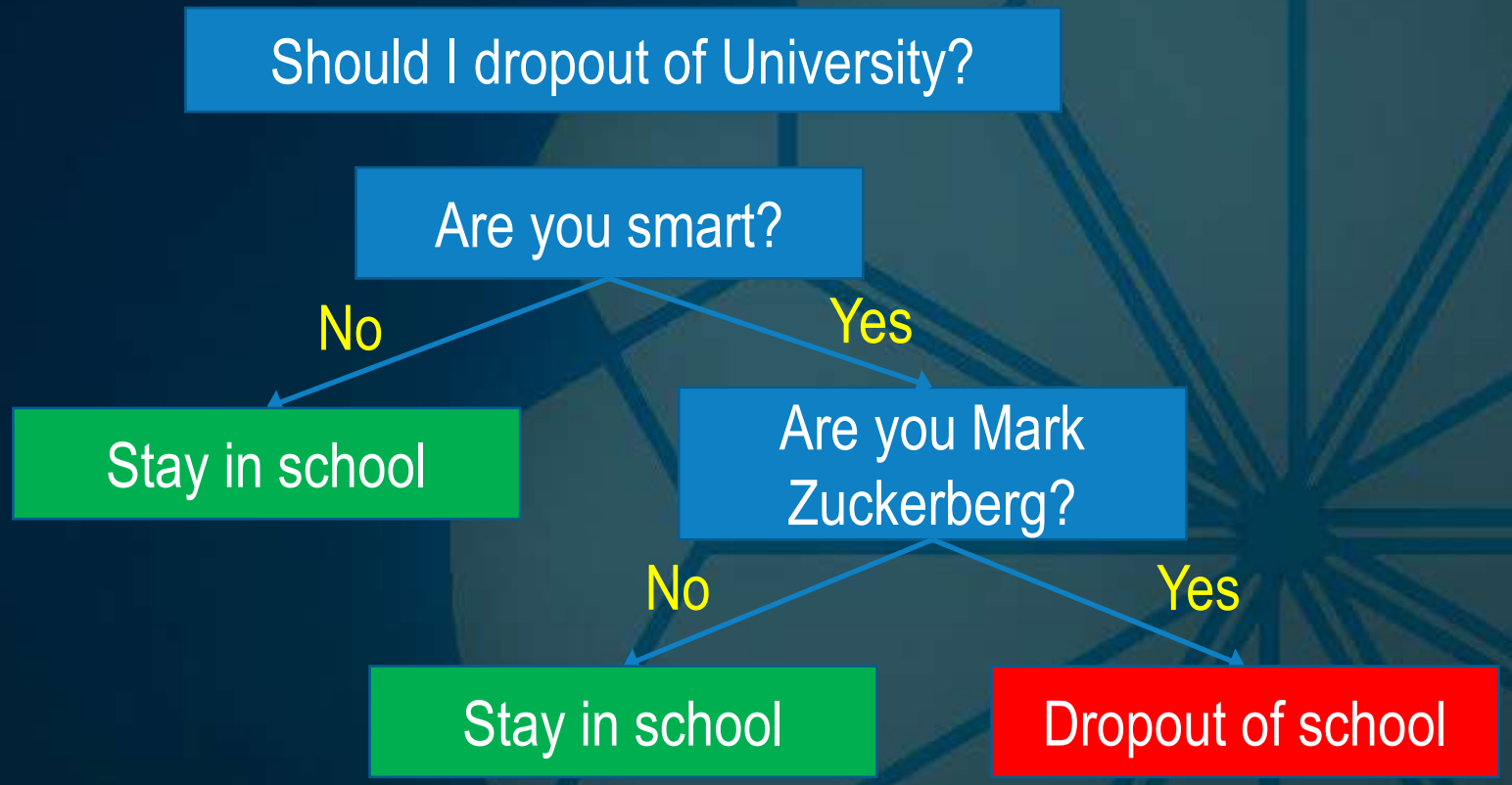
K-nearest neighbors

- Classify new data based on the k-number of training observations the new data was closest to



Decision trees

- Classify new data based on a series of questions with the classification label at the end of the path.



EXtreme Gradient Boosting (XGBoost)

- XGBoost is an ensemble, boosted, decision tree-based model.
- **Ensemble**: Results of multiple decisions trees averaged into 1 result
- **Boosted**: Each additional decision tree is designed to correct misclassified observations

Pros

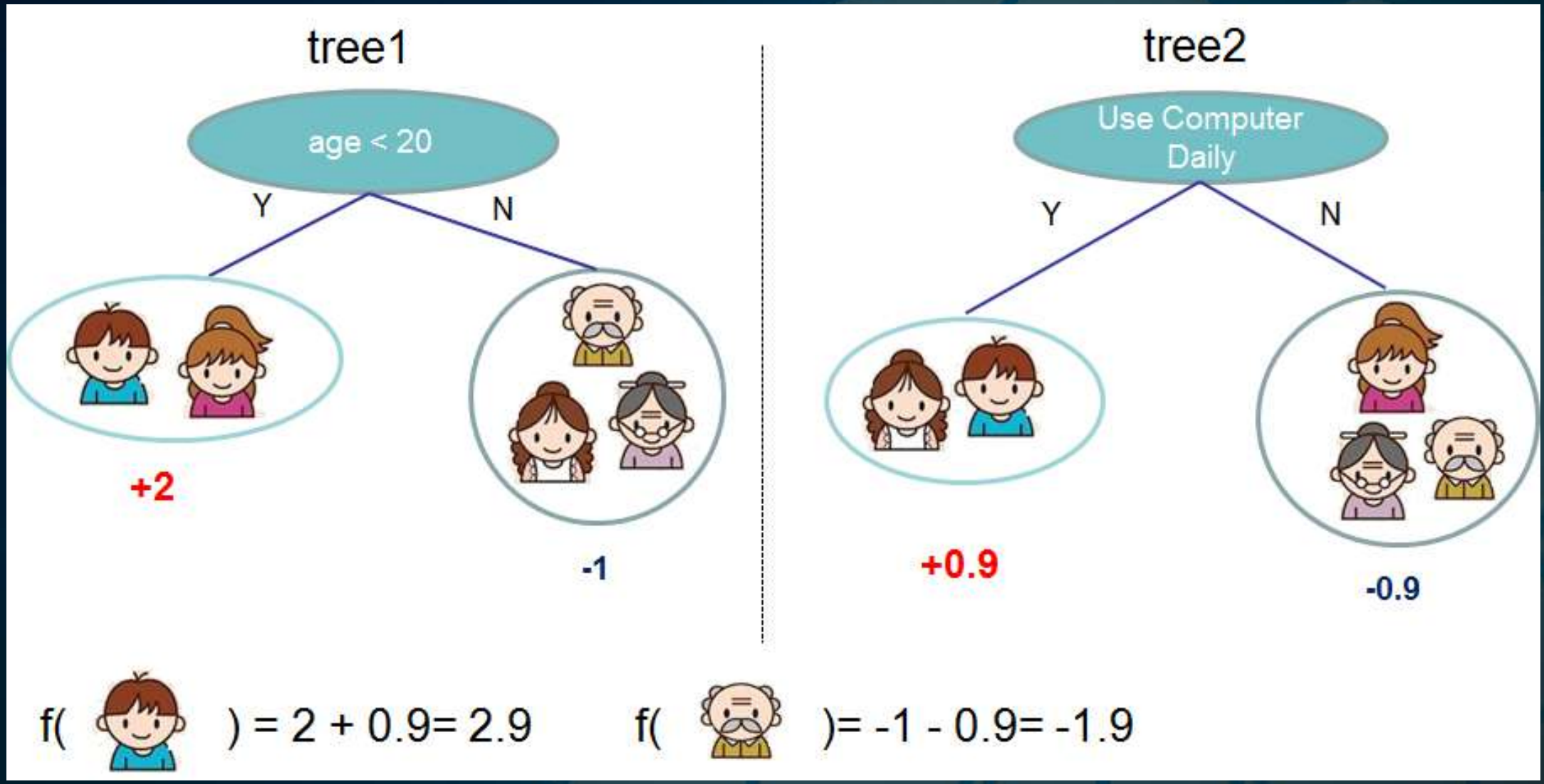
- Very high classification accuracy on tabular data
- Easily handles:
  - Correlative features
  - Features with different scales
  - Missing data
- Regularization helps minimize overfitting
- Very fast training times

Cons

- Models can still overfit training data
- More difficult to explain model

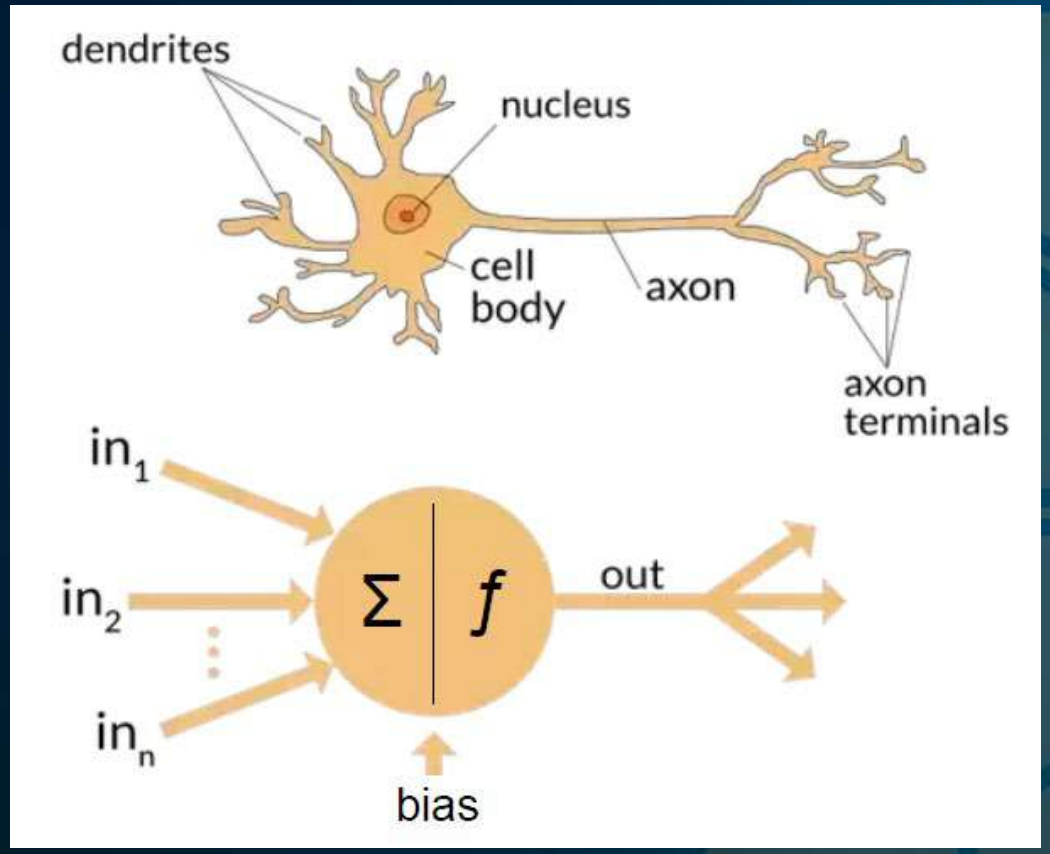
Boosting combines multiple trees to improve prediction

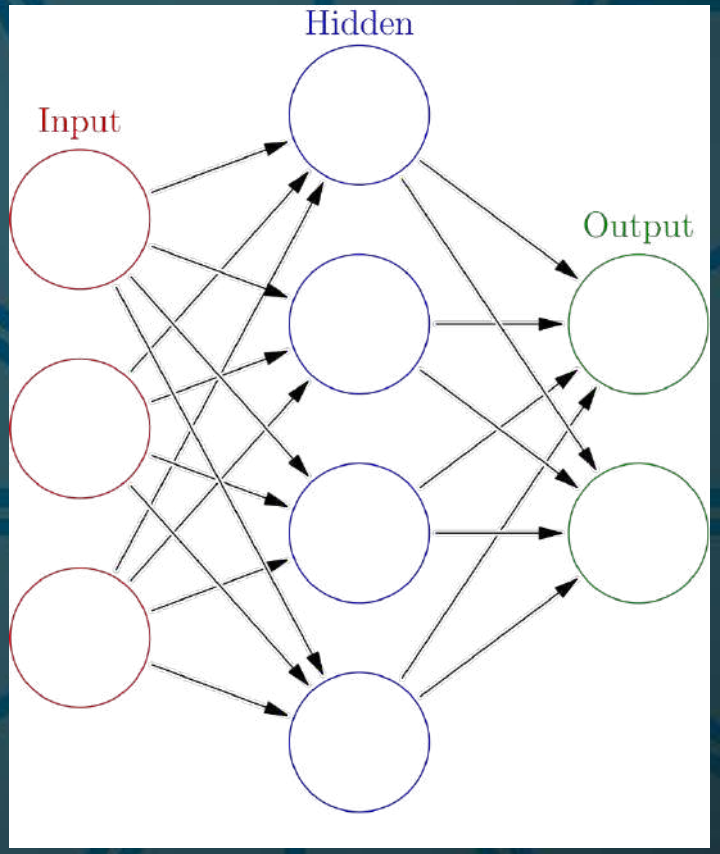
Does this person like video games?



Artificial Neural Network (ANN)

- Inspired by biology, ANN have multiple connected “neurons”. Each neuron has input values multiplied by weights which are summed and modified by a transfer function. This output value is the input value for downstream neurons.

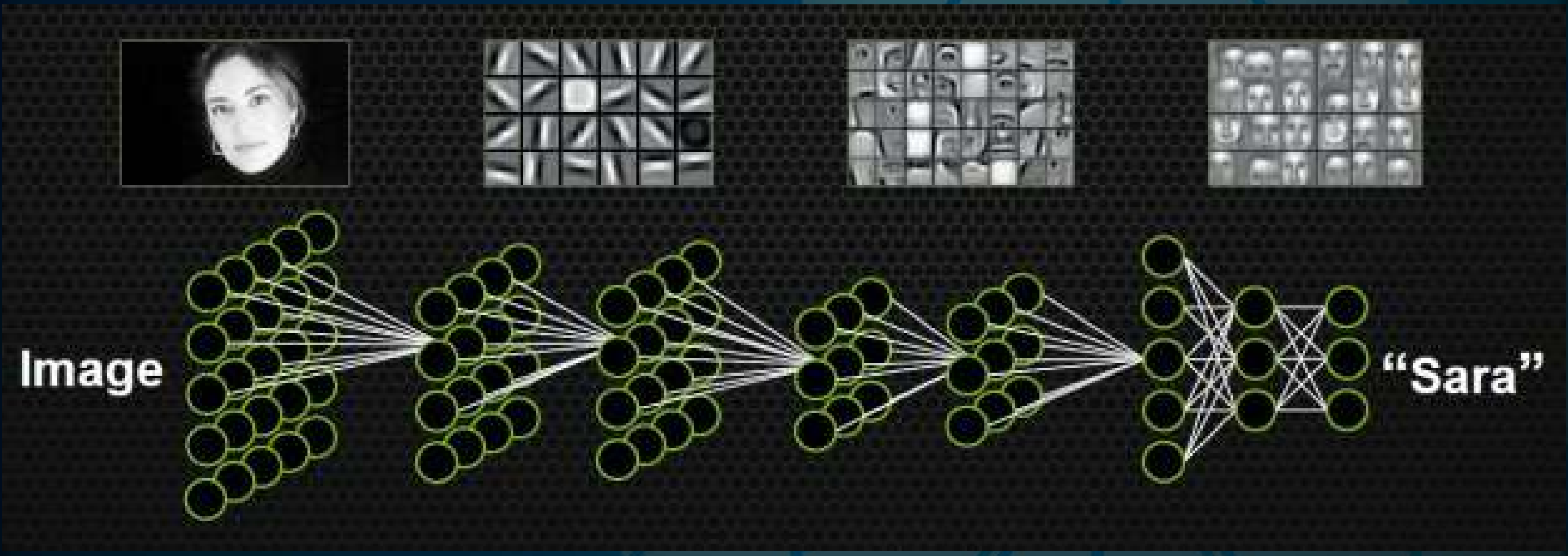




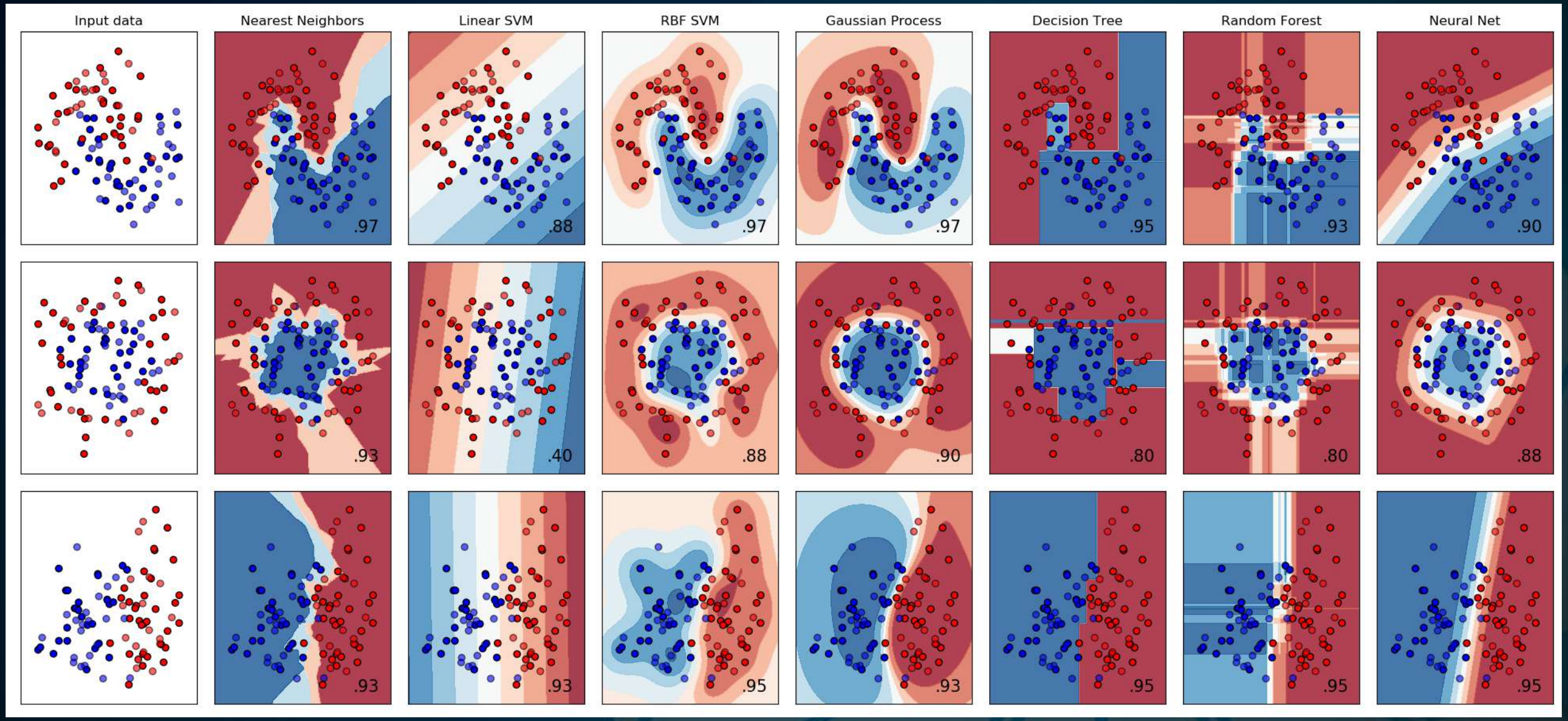
Convolutional Neural Networks (CNN)

Early layers learn edges/colors

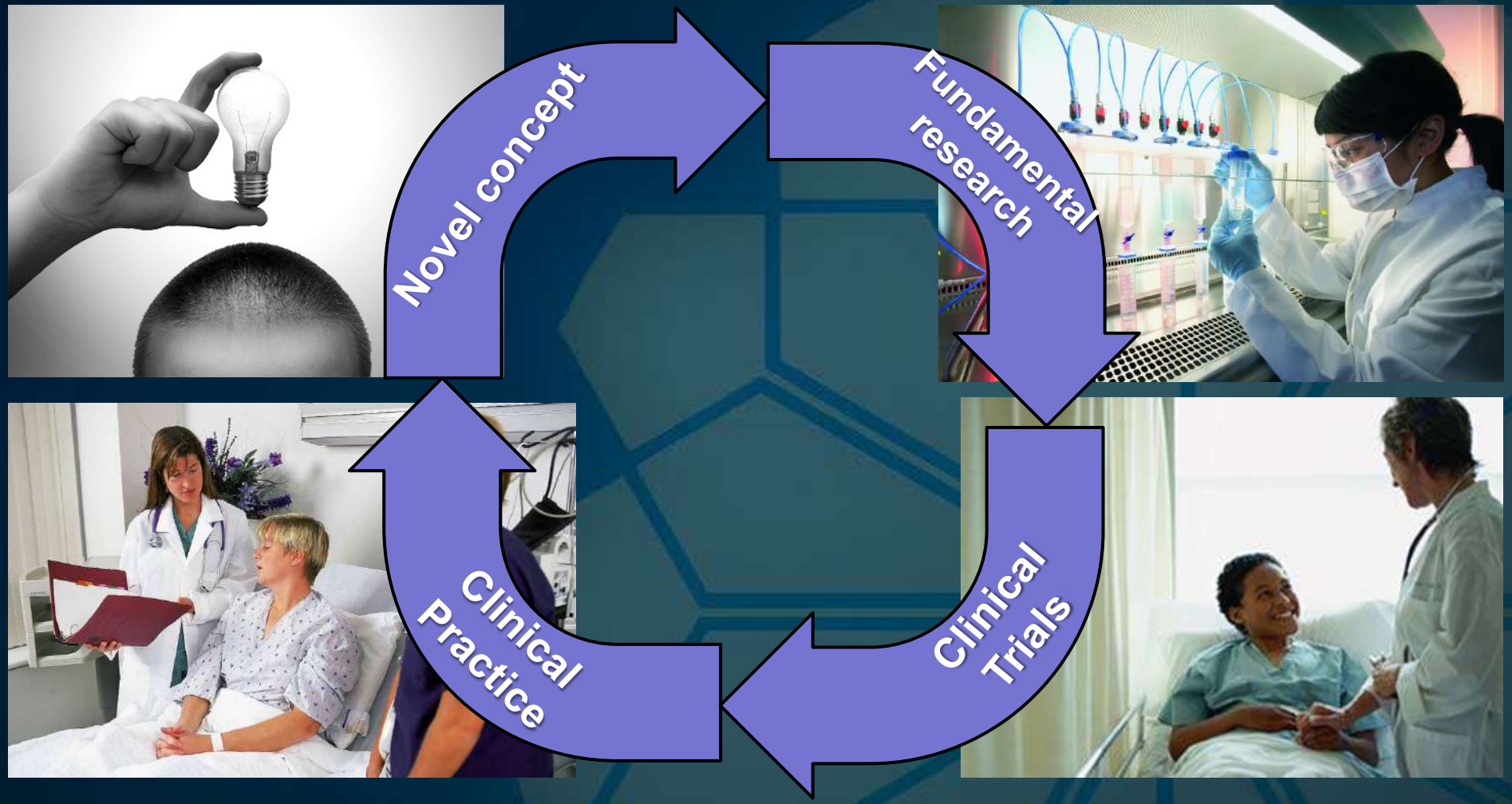
Subsequent layers combine earlier layers to identify useful shapes for classification



Performance of ML algorithms depends on the data



How can we ultimately make an impact on patients?

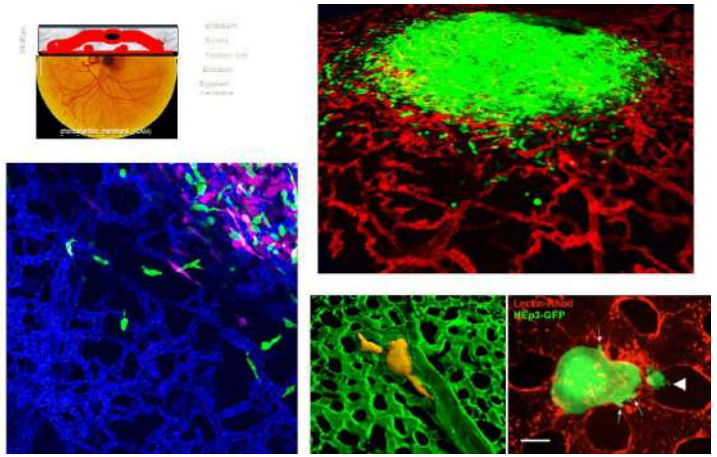


Our clinical translation ecosystem

Discovery Research

The Lewis Laboratory

- Mechanisms of metastasis
- Novel diagnostics
- Nanotechnology for imaging and therapy



Translational Research

ALBERTA PROSTATE CANCER RESEARCH INITIATIVE
knowledge | action | impact

- Founded in 2013 by a provincial multi-disciplinary team of prostate cancer scientists, physicians, patients, healthcare employees and an international collaborative network
- Facilitate translational prostate cancer research
- Accelerate the translation of research from the laboratory to the clinic

Commercialization

NANOSTICS PRECISION HEALTH

entos Pharmaceuticals

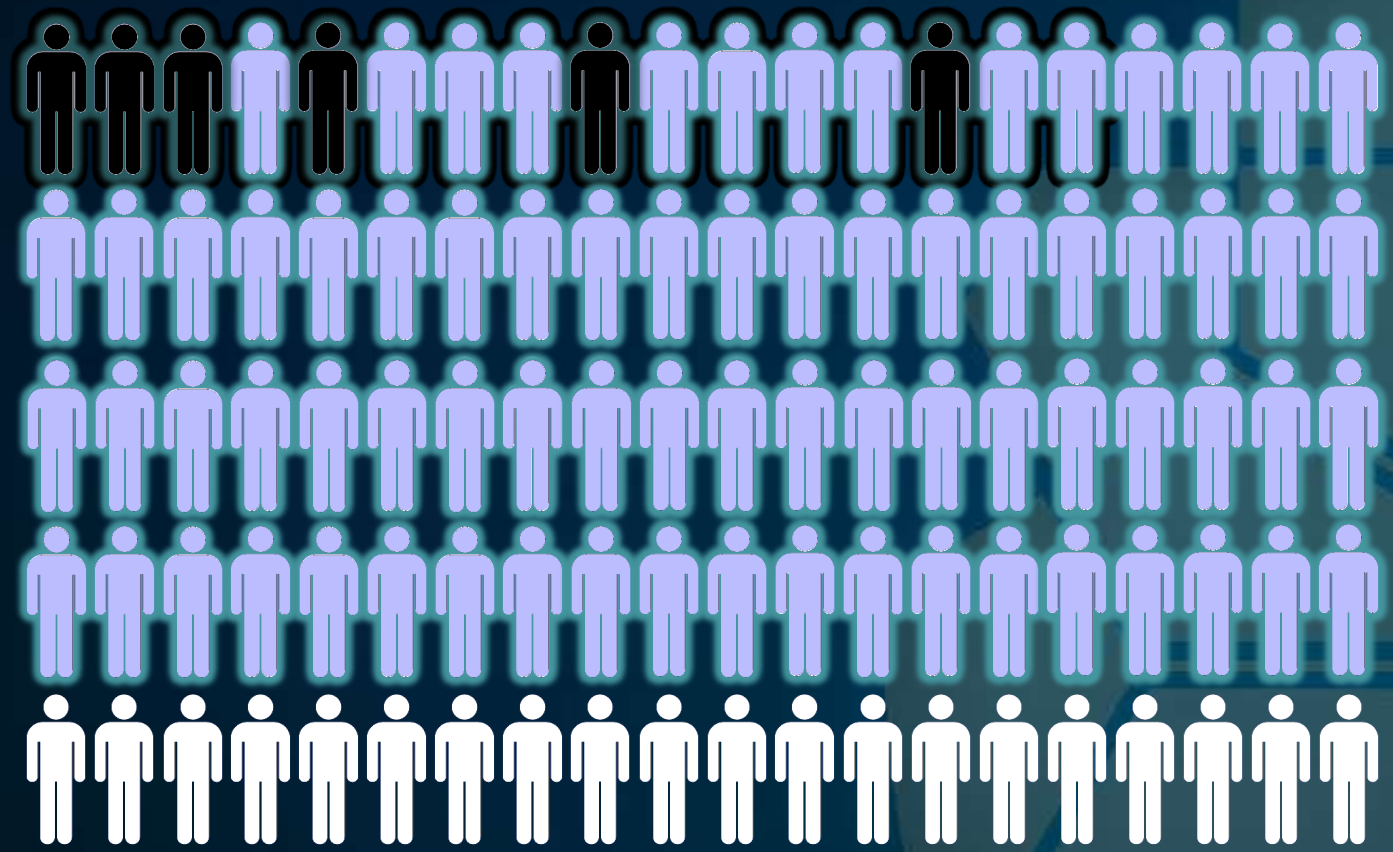
OncoSenX™

AEGIS LIFE, INC.™

OISÍN BIOTECHNOLOGIES®

Prostate cancer is the most commonly diagnosed cancer in men

36% of newly diagnosed cancers, and 10% of all cancer deaths in men



Out of every 100 men...

16 will be diagnosed with prostate cancer in their lifetime

In reality, up to 80 will have prostate cancer by age 70

And 3 will die from it.

But which 3 ?

The deadliest aspect of cancer

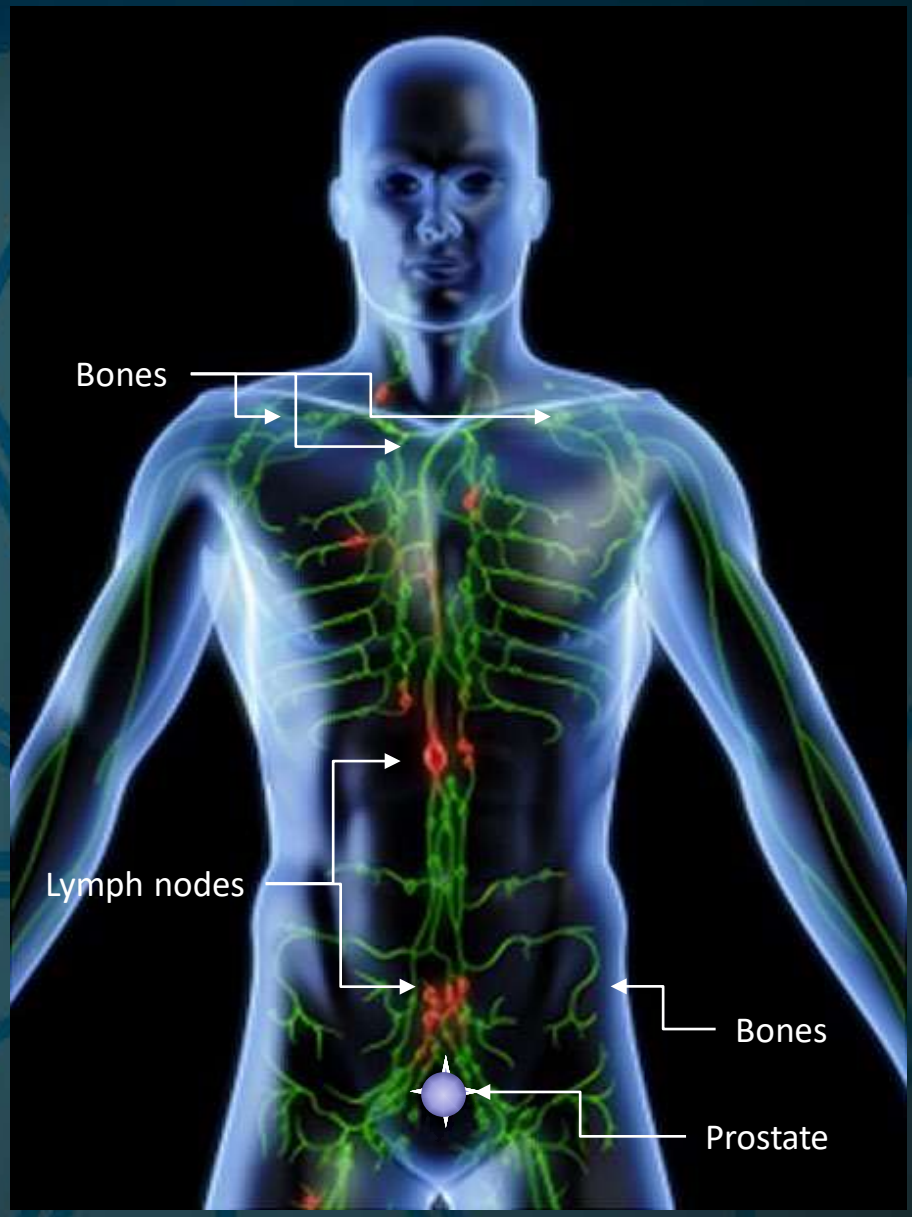
The deadliest aspect of prostate cancer is its spread, or metastasis

In North America, the average 5 year survival rate for localized prostate cancer is 100%

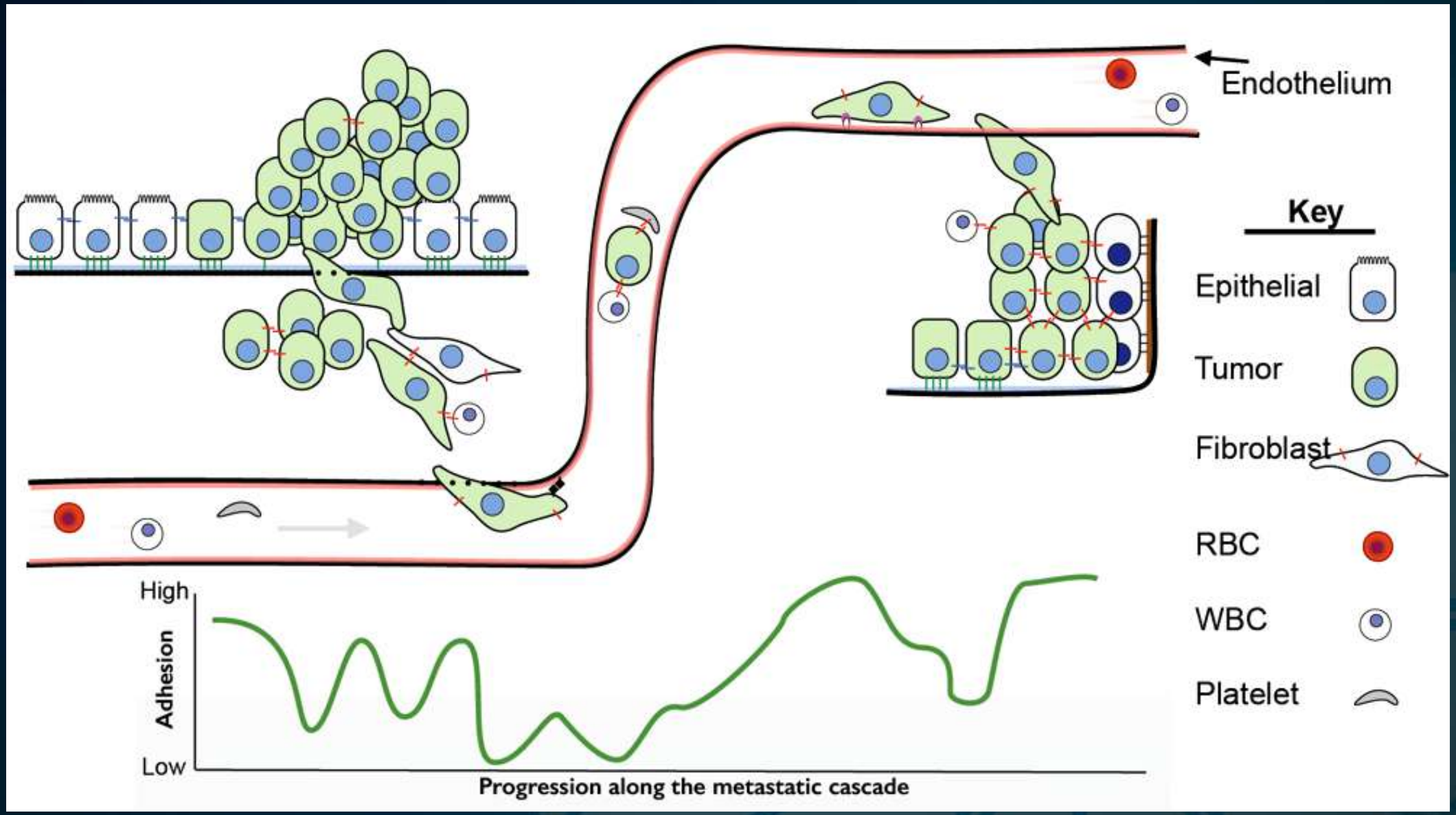
For metastatic cancer, it is less than 30%

Current diagnosis tools do not predict whether metastasis will occur

Current treatments do not prevent or cure metastasis



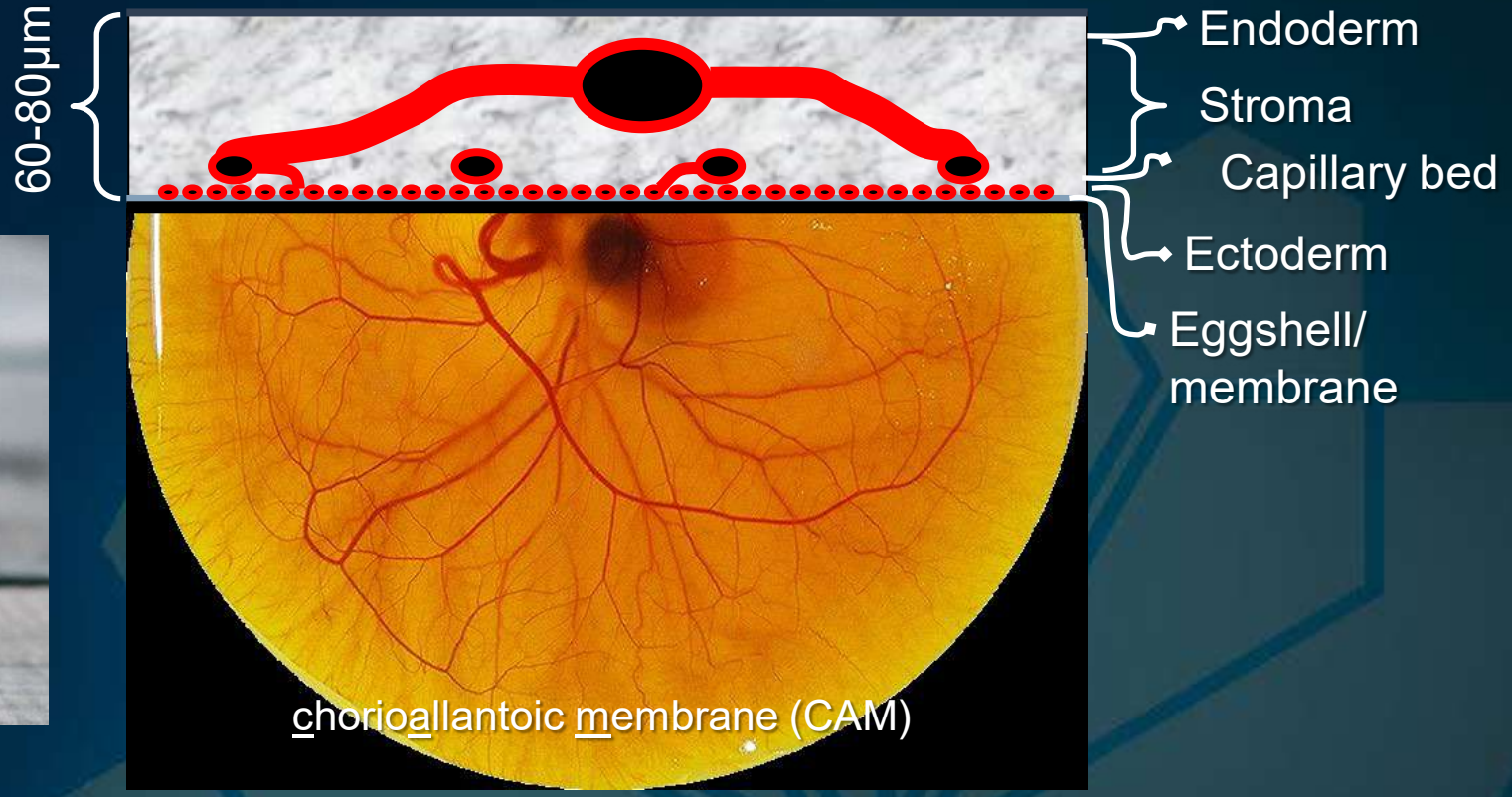
Metastasis is a complex, multi-step process



Palmer et al., Advanced Drug Delivery Reviews 2011

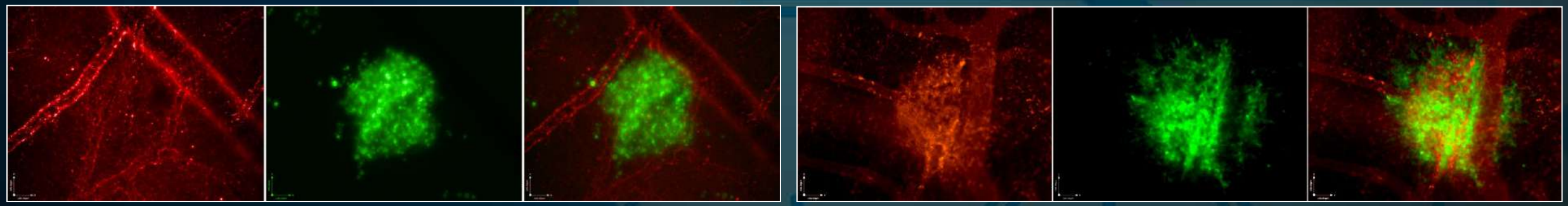
Modeling cancer dynamics in chicken embryos

Leong et al., *Nature Protocols* 2010

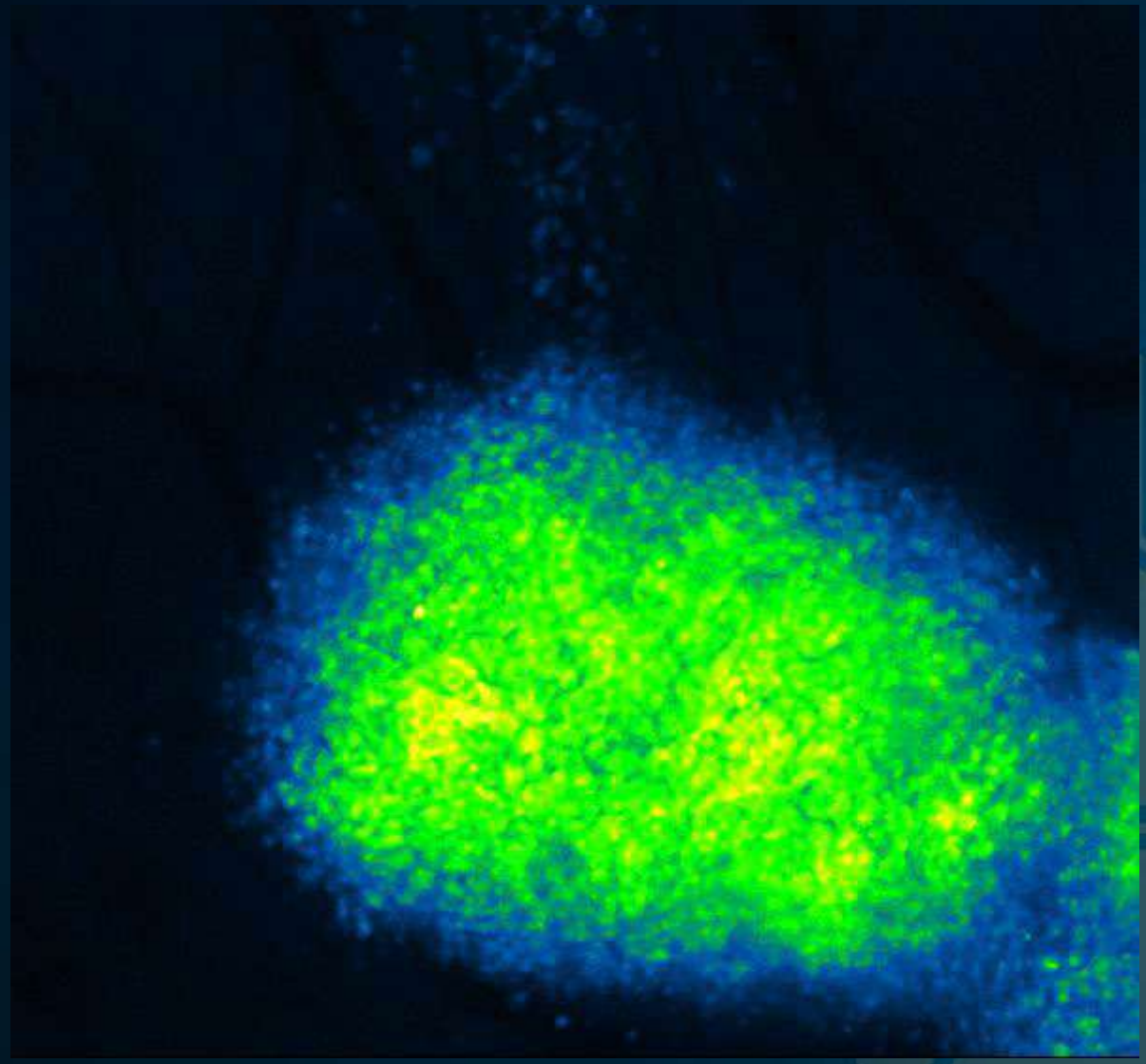


Immediately after injection

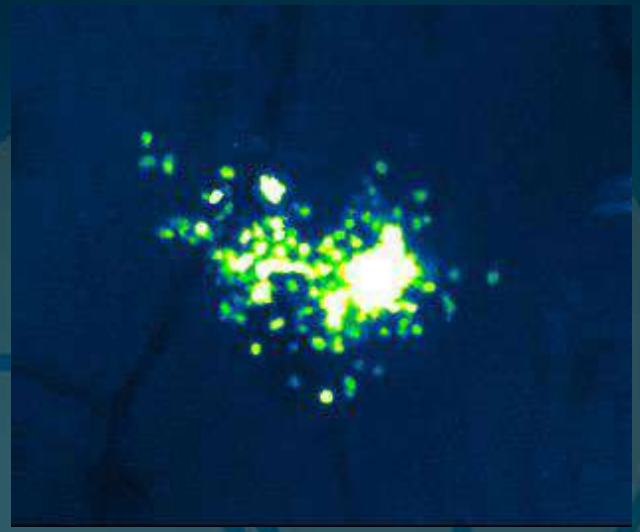
24 hours later



Intravital imaging of tumour growth and metastasis

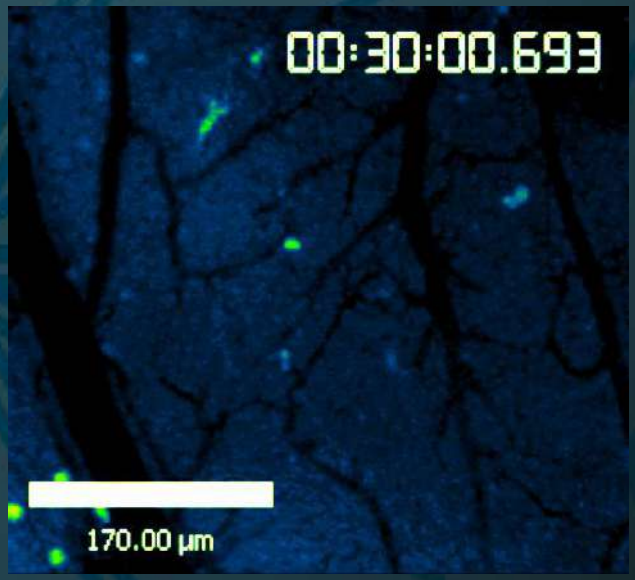


4 mm tumour growing over 4 days

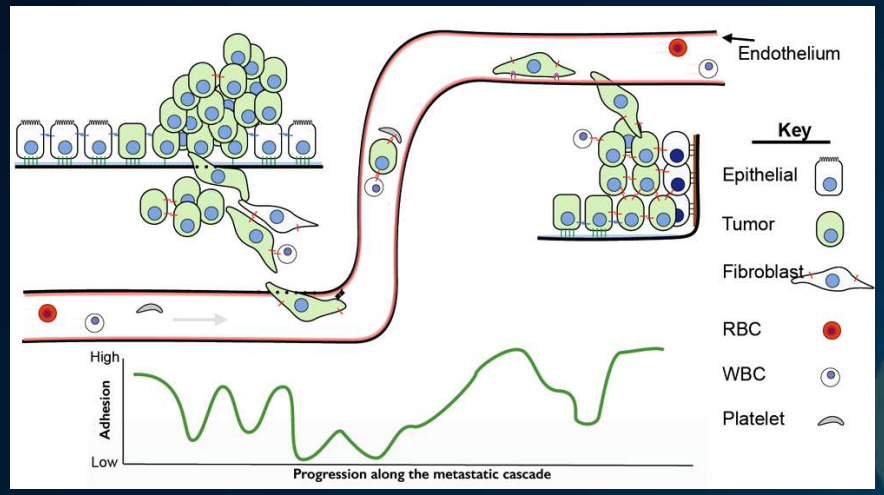


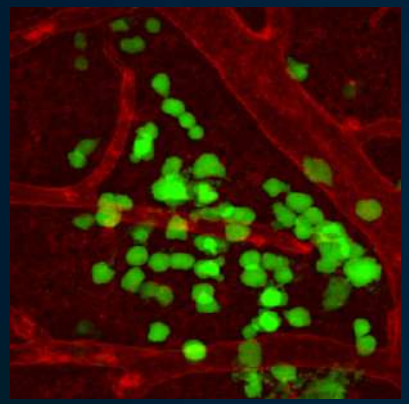
Small metastasis of 200-300 cells (undetectable!)

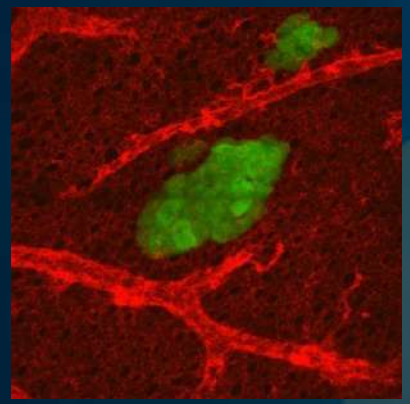
Individual cancer cells (completely undetectable!)



We have identified biomarkers associated with cancer metastasis







Biomarkers related to metastasis were discovered using advanced microscopy approaches

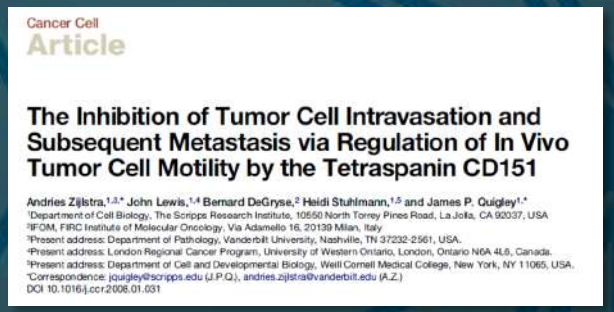

nature COMMUNICATIONS

ARTICLE

Quantitative in vivo whole genome motility screen reveals novel therapeutic targets to block cancer metastasis


Cell Reports
Article

Invadopodia Are Required for Cancer Cell Extravasation and Are a Therapeutic Target for Metastasis

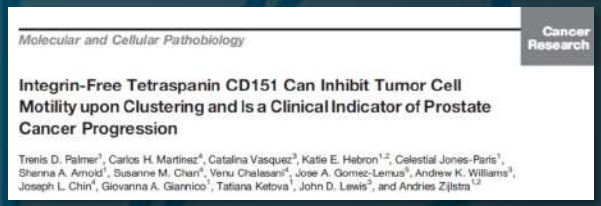

Cancer Cell
Article

The Inhibition of Tumor Cell Intravasation and Subsequent Metastasis via Regulation of In Vivo Tumor Cell Motility by the Tetraspanin CD151


nature COMMUNICATIONS

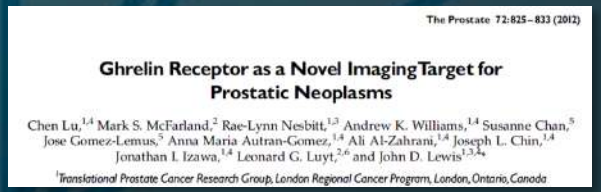
ARTICLE

LPP is a Src substrate required for invadopodia formation and efficient breast cancer lung metastasis

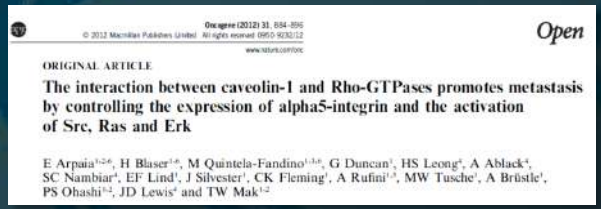

Molecular and Cellular Pathobiology

Cancer Research

Integrin-Free Tetraspanin CD151 Can Inhibit Tumor Cell Motility upon Clustering and Is a Clinical Indicator of Prostate Cancer Progression


The Prostate 72:825–833 (2012)

Ghrelin Receptor as a Novel Imaging Target for Prostatic Neoplasms


Open

ORIGINAL ARTICLE

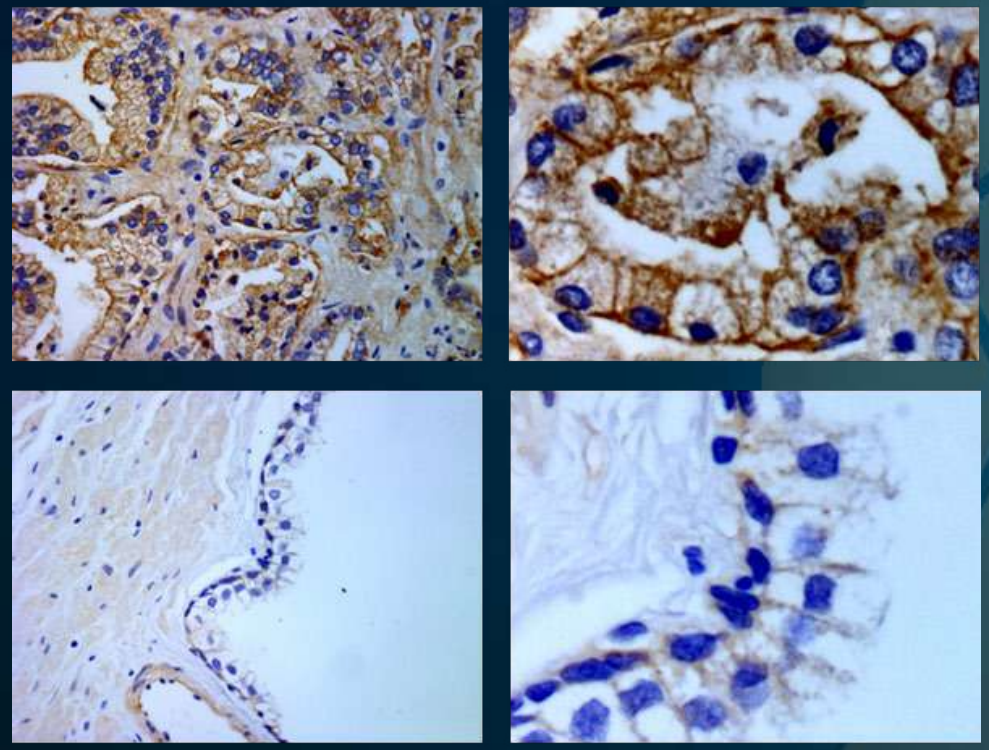
The interaction between caveolin-1 and Rho-GTPases promotes metastasis by controlling the expression of alpha5-integrin and the activation of Src, Ras and Erk

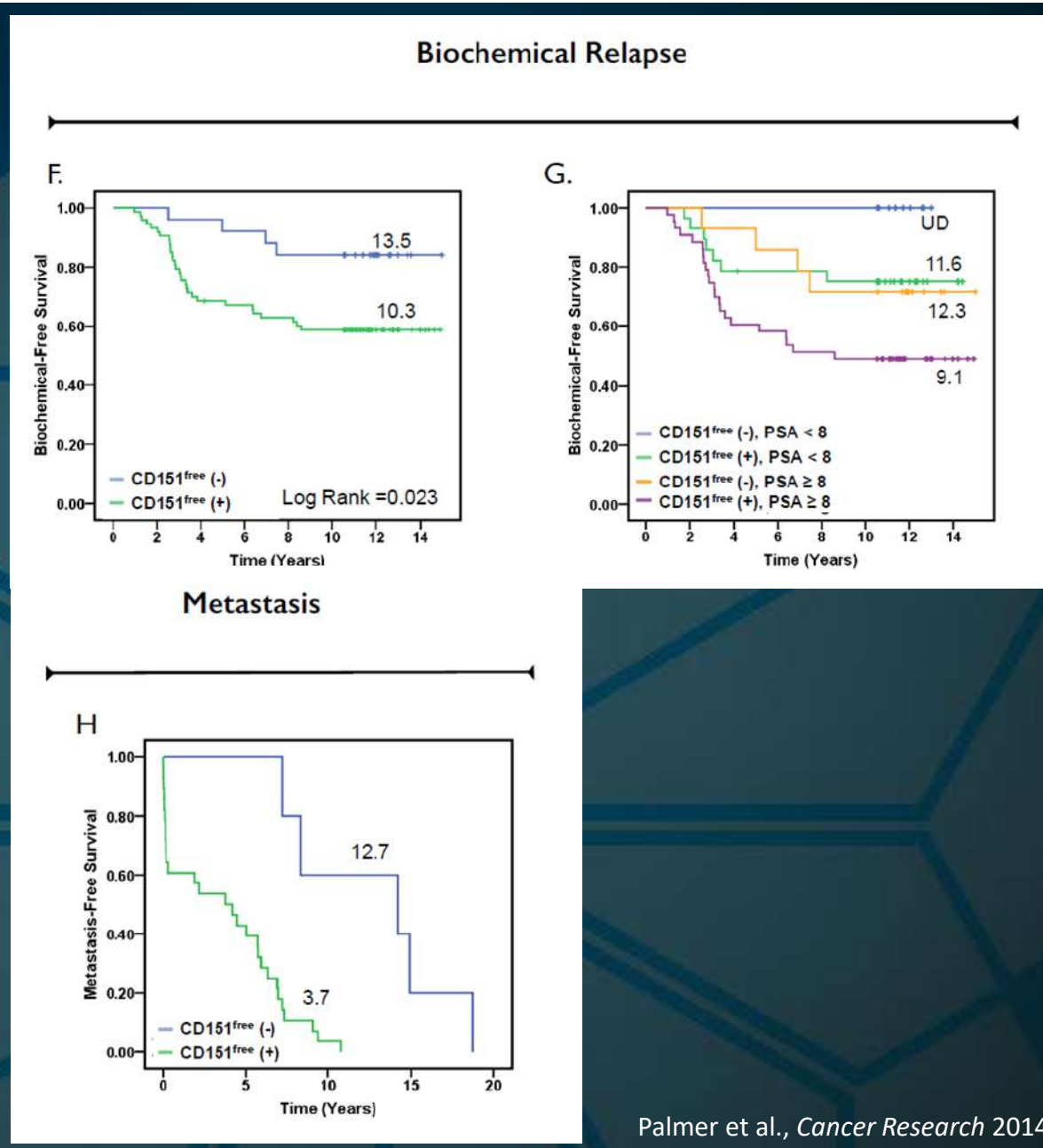
CD151^{free} predicts prostate cancer recurrence and metastasis

138 prostate cancer surgery patients

Follow up: 12.1 years
Recurrence: 34 cases
Metastasis: 38 cases

mAb 1A5





Palmer et al., *Cancer Research* 2014

Detection of disease alone is not sufficient

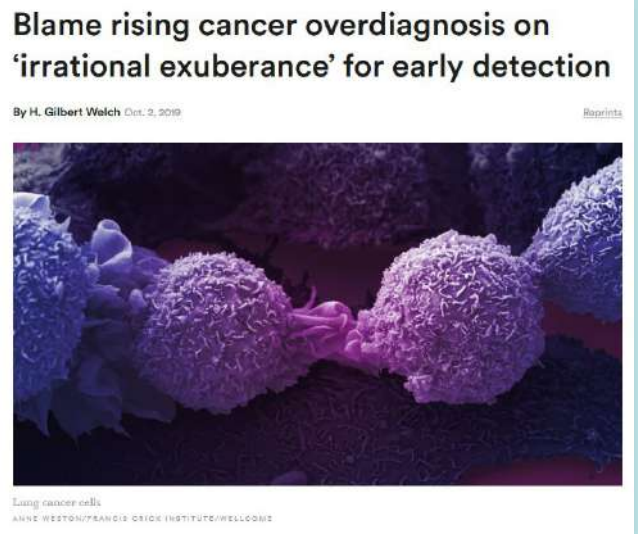
What is the need for intervention?

“Genomic testing is revolutionizing disease screening”

Most new tests identify gene mutations that are simply **associated** with disease, not driving the disease itself

- Will identified mutations be clinically significant?
- Will a potential disease indication ever need intervention?

Gene screening exacerbates the problem of over diagnosis without providing clear guidance on the need for intervention, which dramatically increases over-treatment, reduces patient quality-of-life and increases healthcare costs


Blame rising cancer overdiagnosis on ‘irrational exuberance’ for early detection

By H. Gilbert Welch Oct. 2, 2019 Reprints

Lung cancer cells

ANNE WESTON/FRANCIS CRICK INSTITUTE/WELLCOME


Future Medicine Ltd
Personalized Medicine
Volume 15, Issue 5, September 2018, Pages 343-346
https://doi.org/10.2217/pme-2018-0041

Editorial

Personalized Medicine

Too much of a good thing? Overdiagnosis, or overestimating risk in preventive genomic screening

Karen M Meagher[1] and Jonathan S Berg[2]

[1] [1]Department of Social Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7240, USA
[2] [2]Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599-7264, USA
*Author for correspondence: +1 919 966 7043; jonathan_berg@med.unc.edu


Bulliard and Chiolero *Public Health Reviews* (2015) 36:8
DOI 10.1186/s40985-015-0012-1

Public Health Reviews

COMMENTARY Open Access

Screening and overdiagnosis: public health implications

Jean-Luc Bulliard[1*] and Arnaud Chiolero[1,2]

ALBERTA
PROSTATE CANCER
RESEARCH INITIATIVE
knowledge | action | impact

OVERVIEW of APCaRI

The Alberta Prostate Cancer Research Initiative (APCaRI) is a multi-disciplinary team of:

- Prostate cancer researchers
- Physicians & Clinical research scientists
- Patients
- Industry partners and
- Not for profit networks
- Hospitals & Diagnostic labs

We are working together!

Researchers and Clinicians to positively impact the outcomes and quality of life of those living with prostate cancer by accelerating the translation of new research ideas from the laboratory to the clinic.

ALBERTA PROSTATE CANCER RESEARCH INITIATIVE
knowledge | action | impact

Our network of clinical teams

UNIVERSITY OF ALBERTA

The Northern Alberta Urology Clinic

+ a team of Urologists, MOAs and Clinic Administrator… all committed to the success of the study

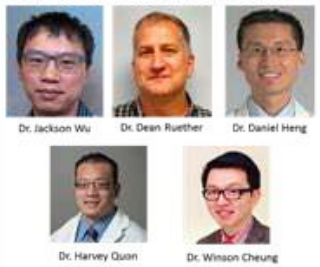
Northern Alberta Urology Centre – Edmonton

Cross Cancer Institute – Edmonton

The Cross Cancer Institute

+ a team of Medical and Radiation Oncologists and the CCI laboratory!!

The Tom Baker Cancer Centre



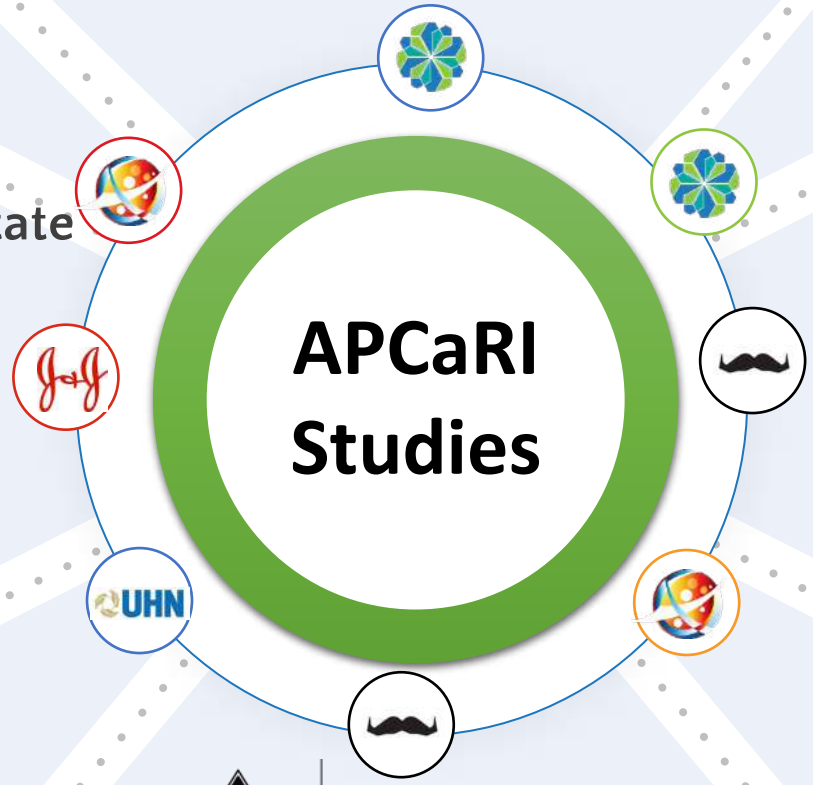
+ a team of Medical and Radiation Oncologists

Tom Baker Cancer Centre – Calgary

Prostate Cancer Centre – Calgary

The Prostate Cancer Centre

+ a team of Nurses, Lab Assistants, Clinical Research personnel, Urologists, MOAs and PCC Director… all committed to the success of the study!!



APCaRI 06

Bladder Cancer Dx test

Recruitment: Dec 2021

APCaRI 01 & 03

Registry & Biorepository

3875 participants recruited in AB

APCaRI 02 & 04

Single Cell Genomics
Familial Prostate Cancer

18 participants recruited in AB

A Study of Men With Advanced Prostate Cancer in Canada (GURC)

Document the course of advanced prostate cancer: disease progression, real-world treatment, and patient management

30 participants recruited in ED

TRUENTH A MOVEMBER INITIATIVE | Prostate Cancer Canada | TRUENTH | GLOBAL REGISTRY PROSTATE CANCER OUTCOMES

Improve outcomes and quality of life of men with prostate cancer and their families.

4978 participants recruited in AB

Metformin Active Surveillance Trial (MAST) Study

Effect of Metformin to avoid PCa Progression

32 participants recruited in ED

TRUENTH | IRONMAN

International Registry for Men with Advanced Prostate Cancer.
EMPRO
50 participants recruited in ED

APCaRI 05

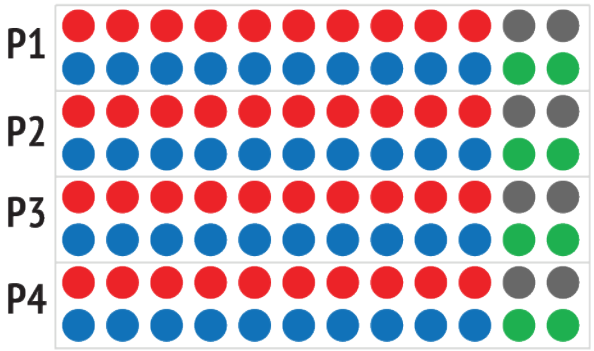
ClarityDX prostate

ClarityDX Validation Study

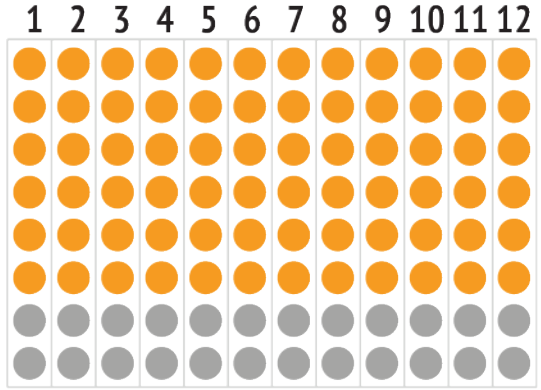
1556 patients recruited
in prospective clinical study in AB

APCaRI Biorepository: a prospective pre-diagnosis cohort in Alberta

Blood: Serum, plasma, buffy coat and red blood cells



Urine and semen



2D Barcode System

Storage at

CANADIAN BIOSAMPLE REPOSITORY

Samples collected and processed – **64 aliquots per patient**

2700
participants enrolled

>150,000
sample aliquots

2 hours
from arm to freezer in 2h -80°C

REDCap
Research Electronic Data Capture

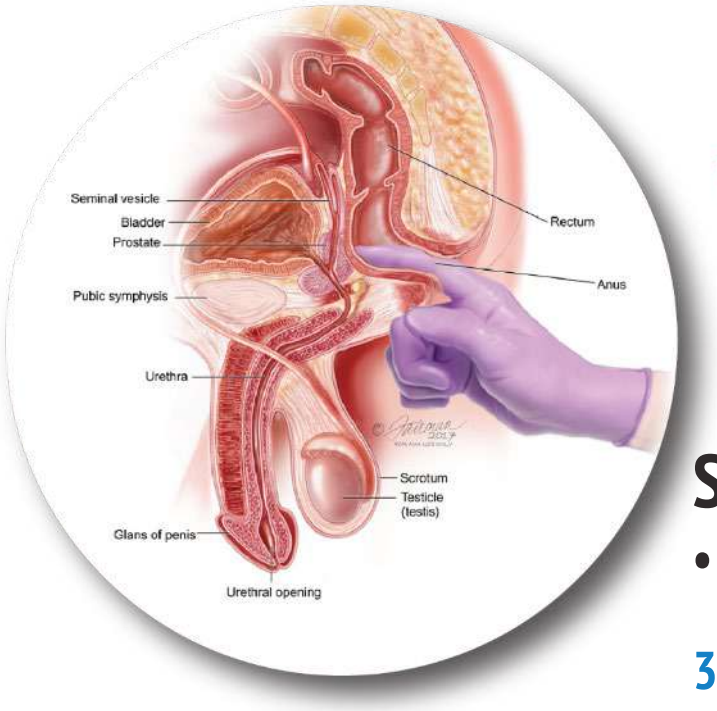
1544 fields
demographic and clinical outcome data

3266
Aliquots distributed for biomarker validation

ALBERTA PROSTATE CANCER RESEARCH INITIATIVE
knowledge | action | impact

http://apcari.ca/our-research/biorepository/

Screening for prostate cancer causes unnecessary harm

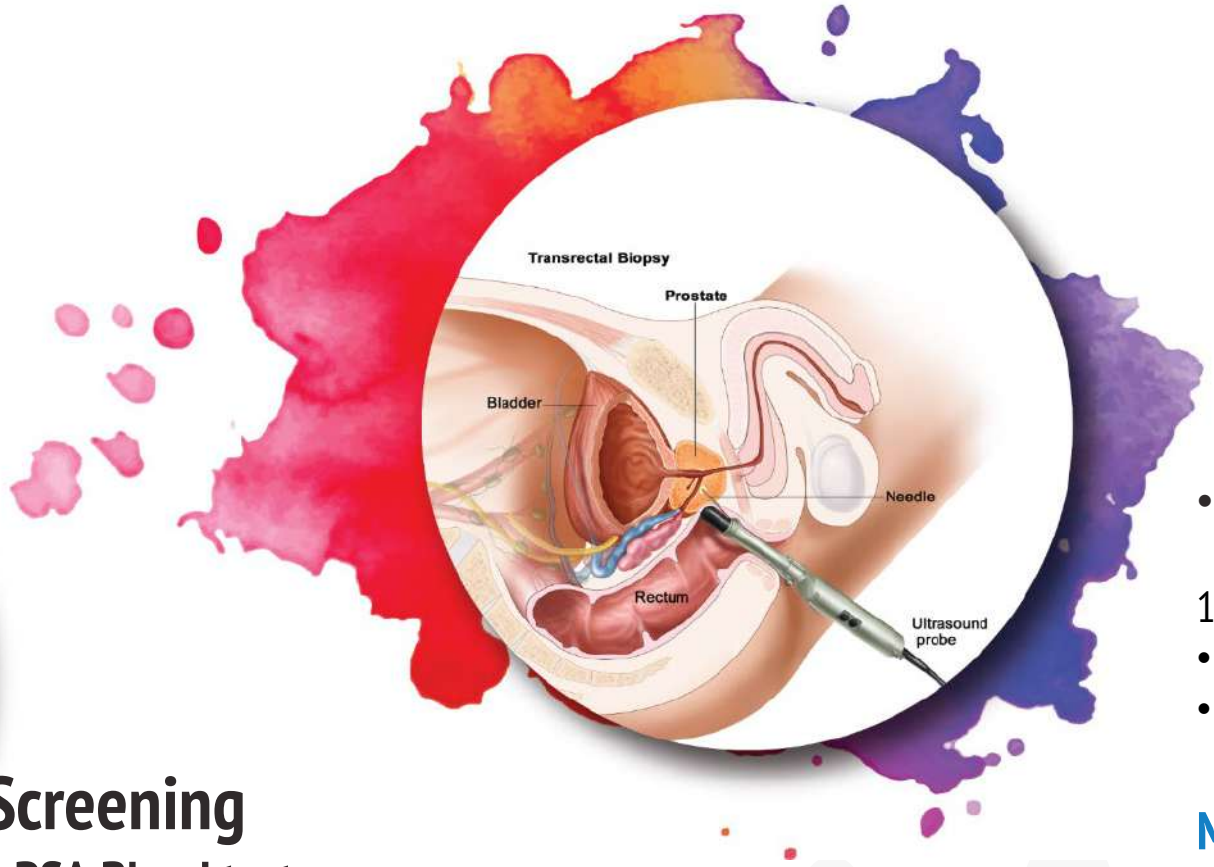


Screening

- PSA Blood test

30 million PSA tests every year in North America

80% of PSA tests are false positives.



Diagnosis

- Biopsy

12 needles through the prostate

- Pain, Discomfort, Infection
- Chance of life-threatening sepsis

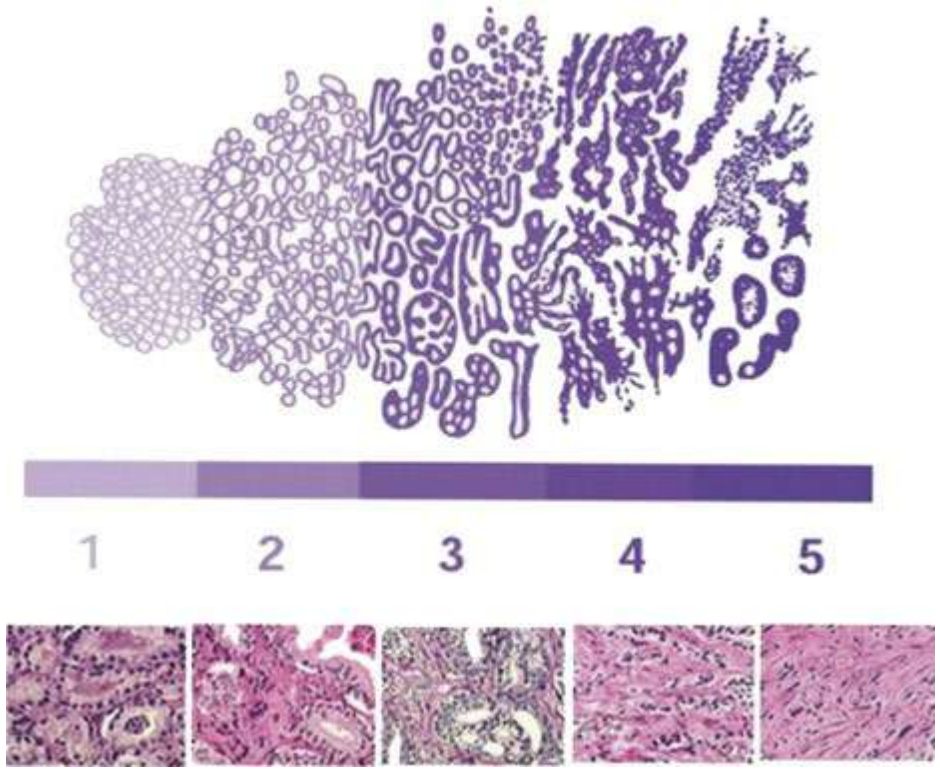
Most are either negative or indolent prostate cancer

Current Dilemmas in Prostate Cancer Diagnosis

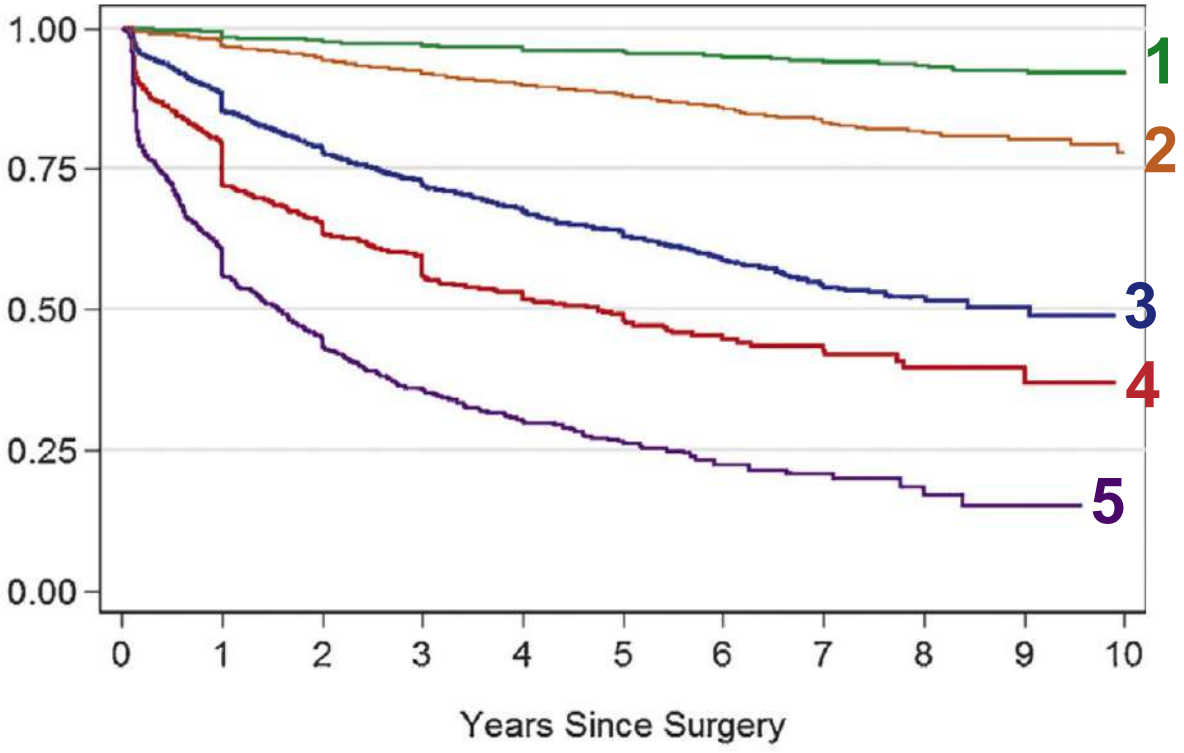
Over detection	• The positive predictive value of PSA (4-10 ng/mL) is low (25-40%)[1] • ~50% to 70% of newly diagnosed cancers are low risk;[1,2] these are unlikely to benefit from early treatment
Unnecessary biopsies	• 65-75% of initial and 10-35% of repeat biopsies are negative[1,3] • Fear of missing significant cancer often results in a repeat biopsy, which is often also negative
Burden of biopsy	• Prostate biopsy may be associated with pain, anxiety and complications[4-6]

1. Hendriks RJ et al. Prostate Cancer Prostatic Dis 2016;20:12-9; 2. Schröder FH et al. N Engl J Med 2009;360:1320-8; 3. Heidenreich A et al. Eur Urol 2008;53:68-80;
4. Loeb S et al. Eur Urol 2013;64:876-92; 5. Wade J et al. J Clin Oncol 2013;31:4235-41; 6. Nam RK et al. J Urol 2010;183:963-8

Prostate cancer is a heterogeneous disease: some forms are lethal, others are not



Men with Gleason Grade Group 3-5 prostate cancers have significantly worse outcomes



Significant need for an accurate, non-invasive test to detect patients that have aggressive prostate cancer

To determine if patients need a biopsy

ClarityDX: combining biomarkers and machine learning

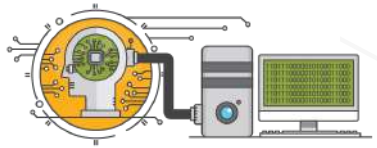
for the prediction of clinically significant PCa

Clinical data

- Age
- Digital rectal exam findings
- Previous negative biopsies

DynaLIFE Dx
Diagnostic Laboratory Services

Machine Learning



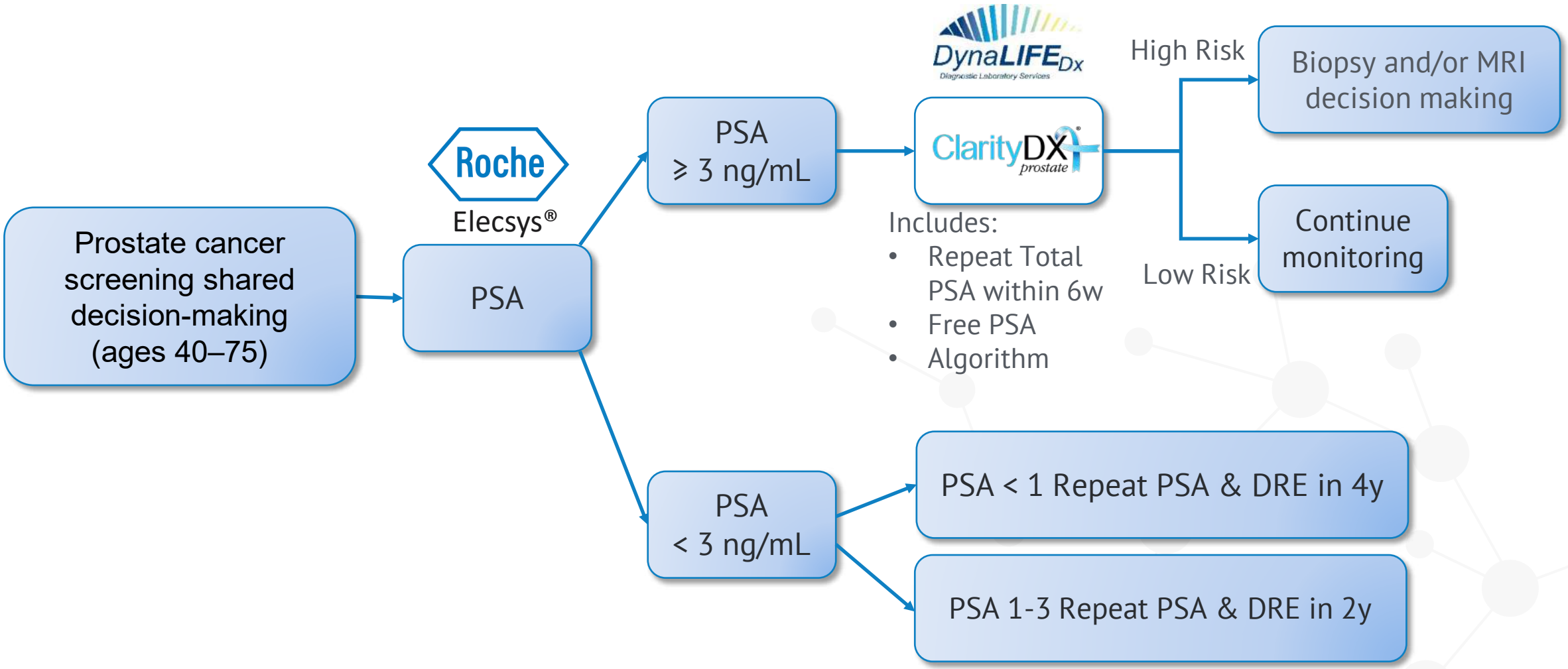
Biomarker data

- Total PSA
- Free PSA

Risk score
Probability of having GG≥2 PCa

ClarityDX® prostate

ClarityDX Prostate® as a reflex test for the PSA test



ClarityDX Prostate Clinical Study Protocol (APCaRI-05)

- 3 Sites Recruiting
  - Kipnes Urology Center
  - Prostate Cancer Center
  - Johns Hopkins University

Male patient, 40-75yo, with abnormal PSA (≥3 ng/mL) & referred for prostate

↓ *n*= 1,584

Identified by urologist / CRC - refer to CRC/research nurse

- Urology clinics
- Pre-biopsy (Diagnostic Radiology Clinics)

↓

- Informed Consent
- Biospecimens are collected
- Intake Surveys (eligibility, medications)
- Demographic and clinical data

↓

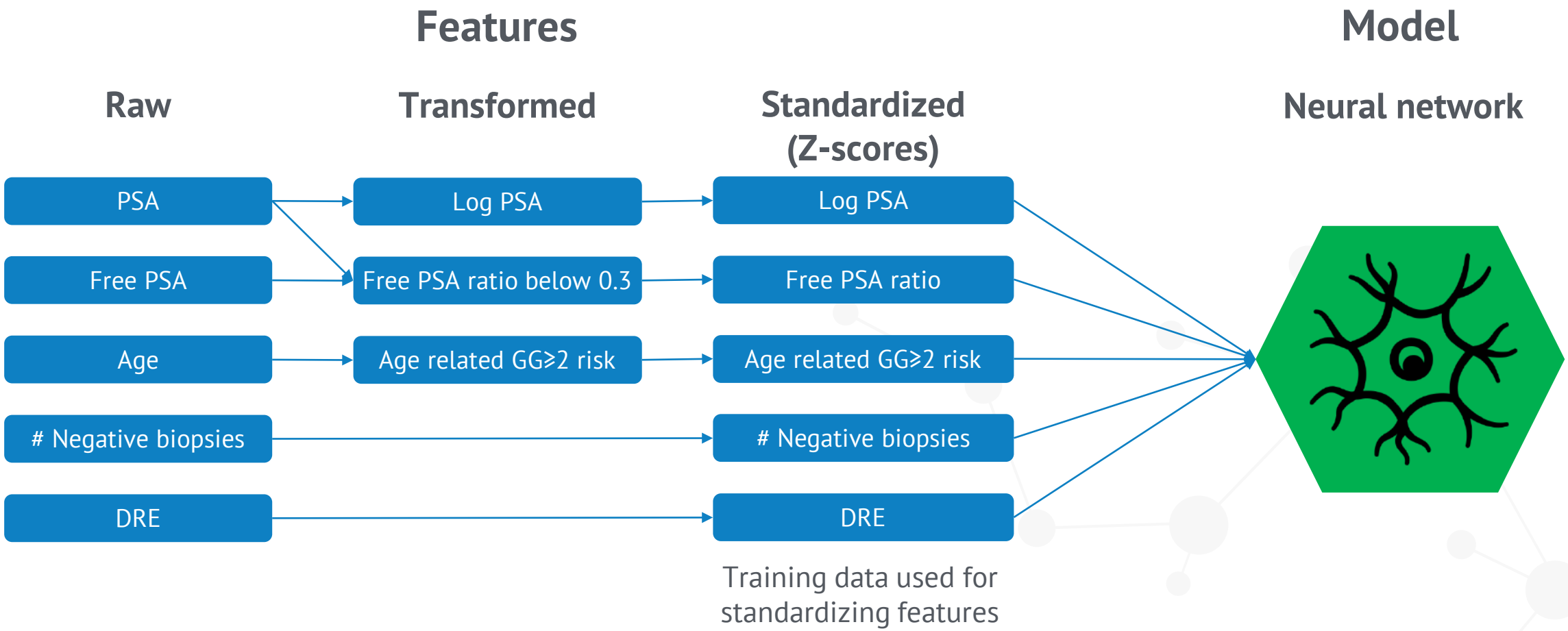
ClarityDX prostate + DynaLIFE Dx Diagnostic Laboratory Services

↓

Biopsy is performed

Performance Characteristics

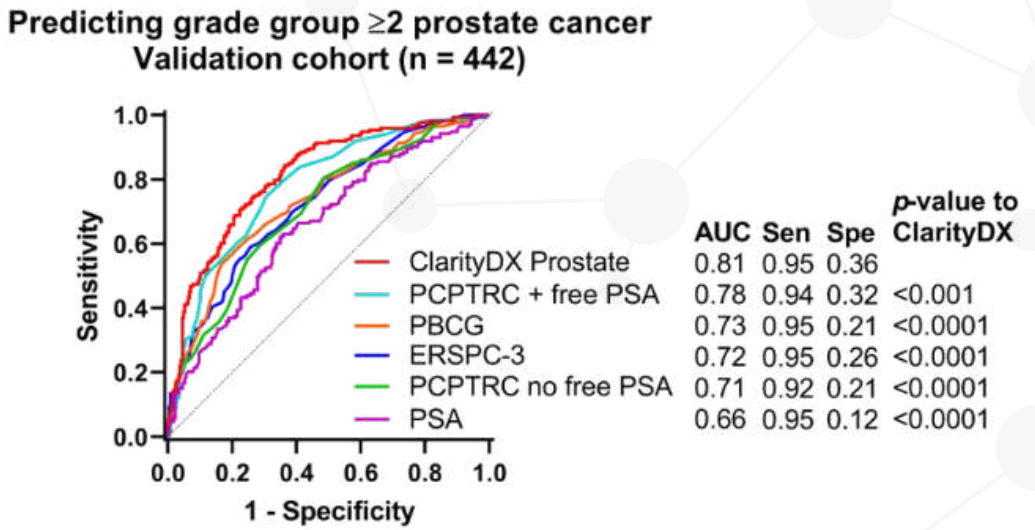
ClarityDX Prostate feature engineering and model



ClarityDX Prostate test has superior clinical performance

	GG≥2 PCa				
	AUC	Sensitivity	Specificity	NPV	PPV
ClarityDX Prostate	0.81	95	36	92	49
PCPTRC	0.71	92	21	81	43
PSA	0.66	95	12	78	41

- Model trained with data from 967 participants from KUC and JHU
- Model evaluated with data from 442 participants from PCC



ClarityDX Prostate is the most accurate model for the prediction of clinically significant PCa

	GG ⩽ 1 PCa	GG ⩾ 2 PCa	p-value	ROC AUC (CI)	Cutoff	Sensitivity, % (CI)	Specificity, % (CI)	PPV, % (CI)	NPV, % (CI)
Patients, n	269 (61%)	173 (39%)							
Prior negative biopsy, n (%)	31 (12%)	3 (1.7%)	<0.0001	0.55 (0.43 - 0.47)		98 (95 - 99)	12 (8.1 - 16)	42 (37 - 47)	91 (76 - 97)
DRE, n (% abnormal)	62 (23%)	82 (47%)	<0.0001	0.63 (0.58 - 0.67)		52 (44 - 59)	74 (67 - 79)	57 (49 - 65)	69 (63 - 75)
Age, yr, median (IQR)	62 (57 - 67)	66 (60 - 69)	<0.0001	0.63 (0.57 - 0.68)	>51.50	95 (91 - 98)	4.8 (2.7 - 7.9)	39 (34 - 44)	59 (36 - 79)
PSA, ng/ml, median (IQR)	6.8 (5.2 - 8.9)	8.5 (6.5 - 12)	<0.0001	0.66 (0.61 - 0.71)	>4.540	95 (90 - 98)	12 (8.4 - 16)	41 (36 - 46)	78 (63 - 89)
PCPTRC, median (IQR)	9.0 (7.0 - 13)	14 (10 - 20)	<0.0001	0.71 (0.66 - 0.75)	>6.500	92 (88 - 96)	21 (17 - 26)	43 (38 - 48)	81 (71 - 89)
ClarityDX Prostate, median (IQR)	**24 (15 - 44)**	**60 (42 - 78)**	**<0.0001**	**0.81 (0.77 - 0.85)**	**>17.80**	**95 (91 - 98)**	**36 (31 - 42)**	**49 (44 - 54)**	**92 (86 - 96)**

ClarityDX Prostate can avoid 36% of unnecessary biopsies

Threshold	GG ⩾2 PCa found (%)	GG ⩾2 PCa missed (%)	Biopsies avoided (%)	Unnecessary biopsies avoided (%)
17.8	95	5	24	36

ClarityDX Prostate can save up to $1.5M per year in Alberta alone

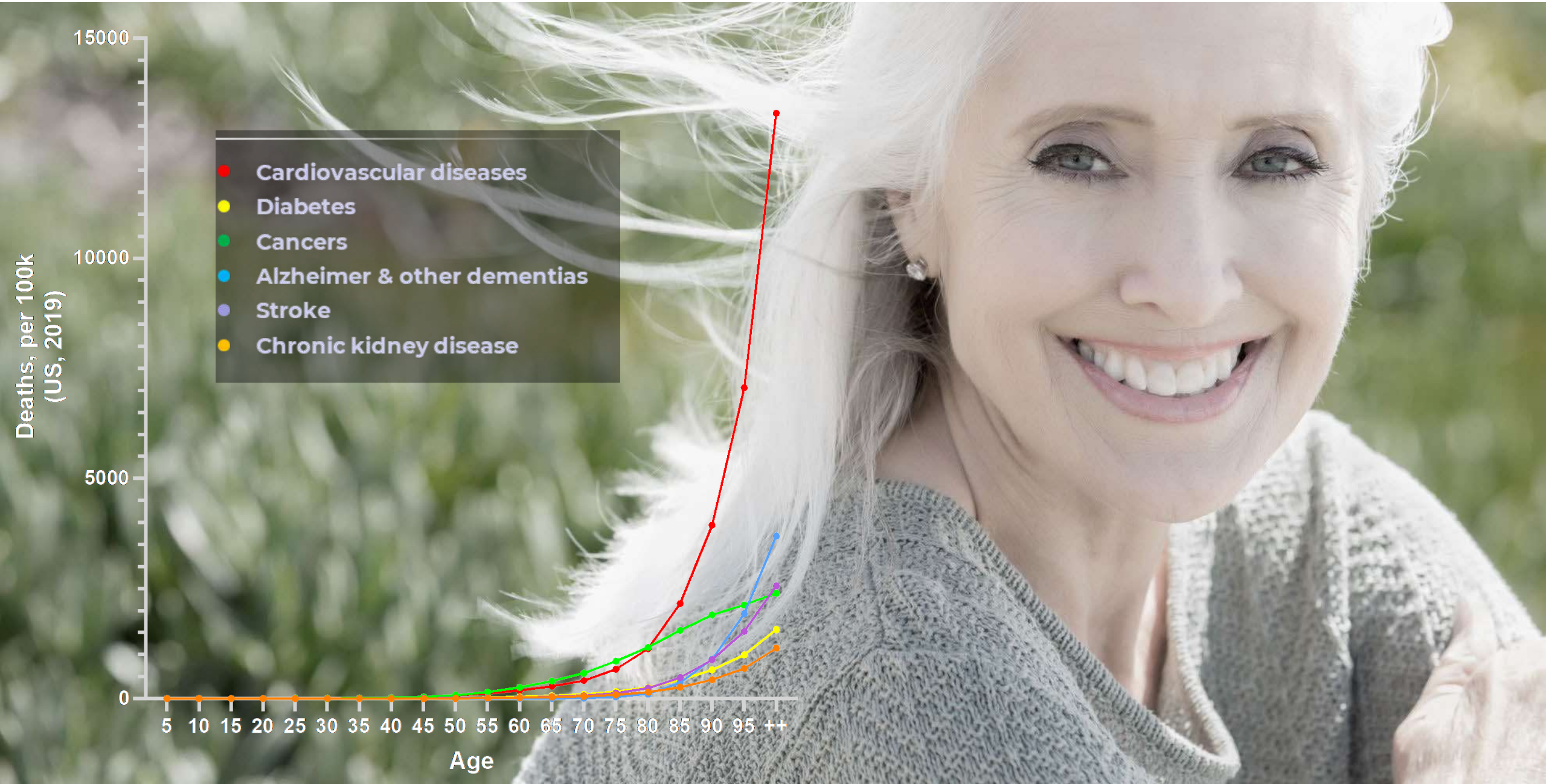
Costs	Avg
Biopsy	$1,148
MRI	$733
Complications	$118
Total	$2,000
Savings Per Elevated PSA	$540
Cost of test	$100
Cost savings per elevated PSA	$440

Building a real-world data asset enabling AI model training and validation

Individuals

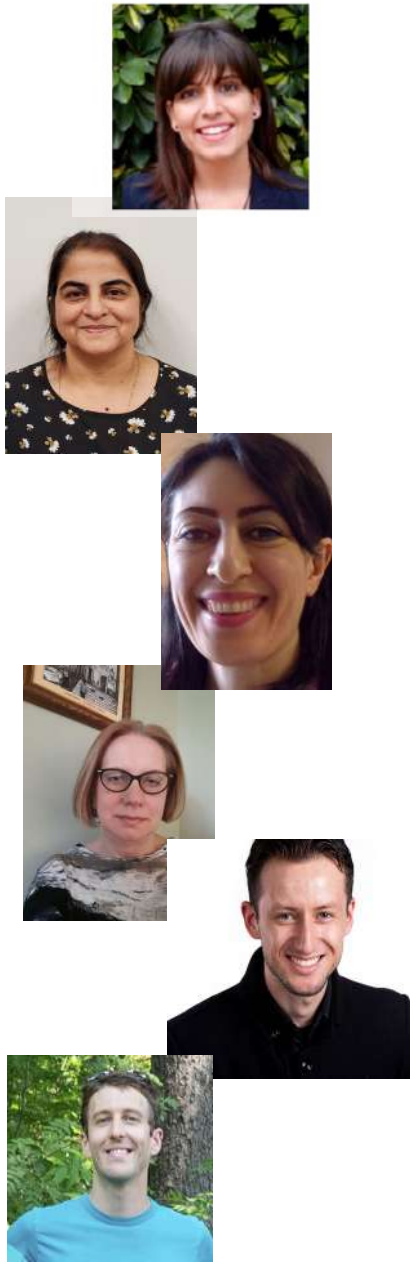
ALBERTA PROSTATE CANCER RESEARCH INITIATIVE	9,000+	Longitudinal patient data: Prostate cancer records
UCLA	1,500	Prostate cancer records
Alberta Health Services	2M	Alberta Cancer Registry: access to all cancer data from Alberta
CanPath Canadian Partnership for Tomorrow's Health	330,000+	Longitudinal patient data: Canadians aged 30-74
MOVEMBER TRUENORTH	40,000+	Longitudinal patient data: men with prostate cancer globally

What's next?


15000
10000
5000
0
Deaths, per 100k (US, 2019)
Cardiovascular diseases
Diabetes
Cancers
Alzheimer & other dementias
Stroke
Chronic kidney disease
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 ++
Age

Acknowledgements

Team	Study
Anais Medina Martin, PhD	APCaRI 01&03 APCaRI 05 MAST
Jay Bhatt, MD	APCaRI 01&03 APCaRI 05 IRONMAN TrueNTH-PCO
Parisa Hamayeli-Mehrabani, PhD	TrueNTH-PCO IRONMAN APCaRI 01&03 GURC
Yuliya Kulchytska, MSc	APCaRI 01&03 IRONMAN TrueNTH-PCO
Diego Ordonez-Pasaje, MD	APCaRI 01&03 APCaRI 05 MAST
Maxwell Uhlich, MSc	APCaRI 01&03 PROSPeCT



Acknowledgements

Alberta Cancer Foundation

Alberta Innovates Health Solutions

Canada Ride for Dad

Bird Dogs For Prostate Cancer Research — Pointing to the cure

Prostate Cancer Canada

Movember Foundation

CanPath — Canadian Partnership for Tomorrow's Health

Johnson & Johnson Innovation

University Hospital Foundation — Supporting innovation and excellence at the University of Alberta Hospital, Mazankowski Alberta Heart Institute, Kaye Edmonton Clinic