



drug discovery AI

Executive summary



Intellomx is a UK-based firm leveraging innovative AI and machine learning to enhance and de-risk drug discovery through precise *in silico* technologies:

- **Biology:** Our USP lies in our capacity to understand the fundamental biology of disease. This insight allows us to pinpoint the best drug targets and choose the drug candidates most likely to affect these targets.
- **Validation:** Our findings have been confirmed in human biology and plant physiology, leading to the creation of in-house assets and intellectual property for both Intellomx and its collaborators.
- **Data:** Data is no longer a constraint. We harness both public and proprietary information in genomics, proteomics, metabolomics and transcriptomics to discover new targets that influence disease.
- **Swarm Algorithm** Our algorithms invoke multiple, parallel swarms of analyses rather than compute-resource hungry deep-models. Swarms are 90% more efficient, minimize risk of false discovery and are explainable.
- **Collaboration:** Our twin-track business approach features (1) a “fee for service” option, underpinning our internal research discovery initiatives, and (2) shared IP partnerships.
- **The future:** Target ID and docking done, Intellomx is building an integrated model of human disease pathways, captured in the form of the Intellomx Digital Twins - heralding the future in drug discovery.

Intellomx leads the way in this new era of precision drug discovery, utilizing its proprietary AI/ML to identify optimal drug targets and linking those targets to ranked drug candidates – actionable insights for pharmaceutical collaboration.

Pharma's problem, AI's answer



"A selective high-quality molecule will never become a medicine if it is modulating the wrong target. This is why target selection is the most important decision we make in research."

Mene Pangalos,
former VP, AstraZeneca



"Artificial intelligence is the key to unlocking the vast potential of novel biological targets, transforming the way we discover new drugs. By intelligently mapping the complexity of biology, AI accelerates the identification of promising targets, enhancing the precision and speed of therapeutic breakthroughs."

Demis Hassabis
CEO of Google DeepMind



Solution – Intellomx I³



I ³ -Distiller	<i>Identification of key genes</i>
I ³ -Driver	<i>Identification of key disease drivers</i>
I ³ -Miner	<i>Defining the disease pathway</i>

The Intellomx I³ toolkits model different aspects of the underlying systems biology of disease, using proprietary Artificial Neural Networks and Machine Learning techniques developed in-house.

The primary tools (Distiller, Driver and Miner) model disease pathways at the molecular level whilst our Digital Twin anticipates toxicity liabilities.

I³-Digital Twin *Anticipating tox liabilities*

I³-Pilot *Identification of novel Small Molecules and Biologics*

I³-Precise *Panel optimization for diagnostics*



Intellomx I³: Intuitive, Informed, Intelligence

Solution – Intellomx I³

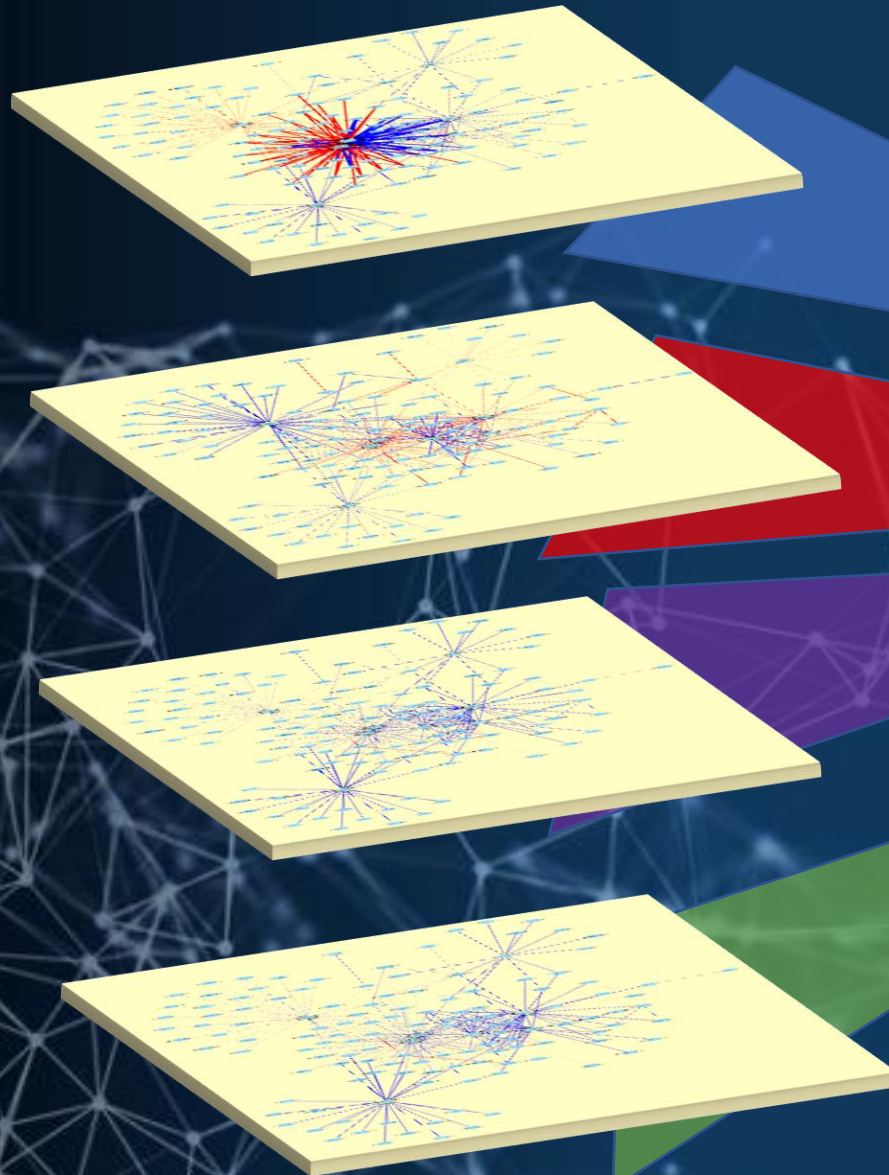


Digital Twin

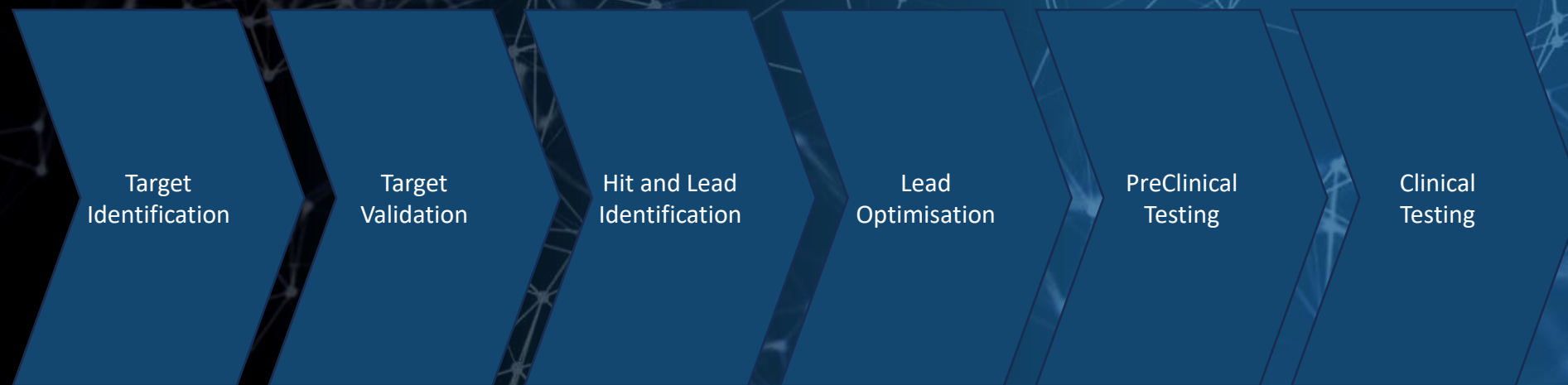
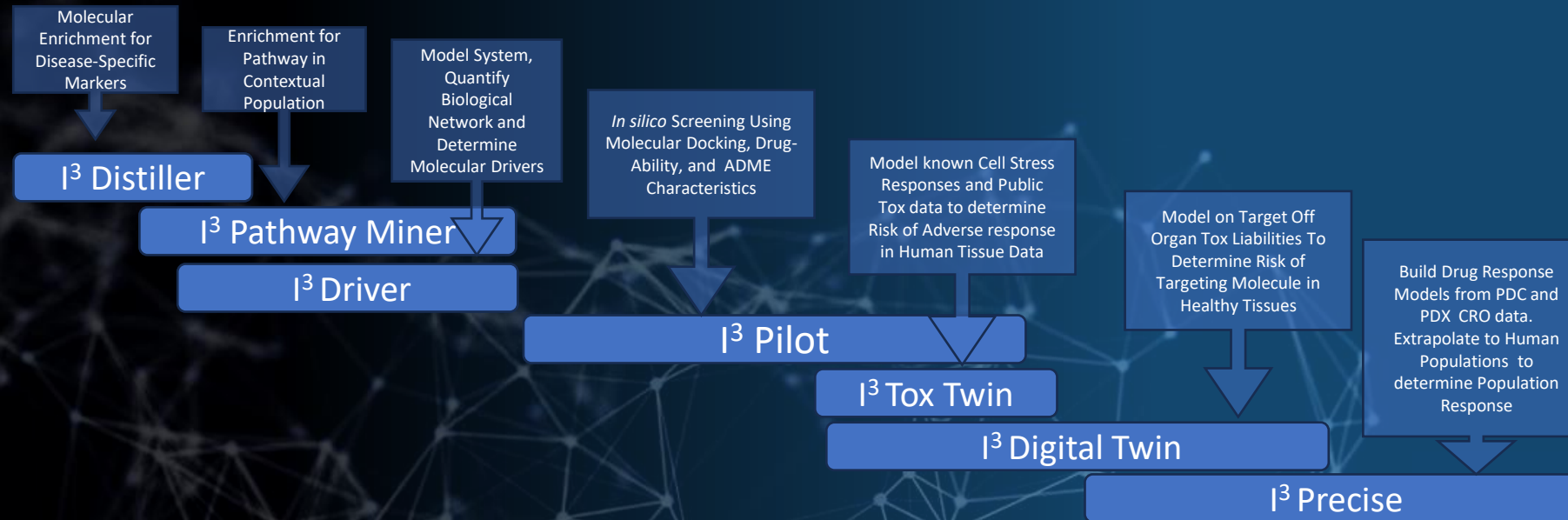
The Intellomx Digital Twin tests drugs in development in an AI-generated human model without risk, giving a clear indication of off-target toxicity.

We can thus prioritise molecules for development, eliminating up to 90% of projects that would fail due to toxicity in later, high-cost stages.

In future, a population of Intellomx Digital Twins will enable *in silico* clinical trials.



Solution: I³ for drug discovery/development



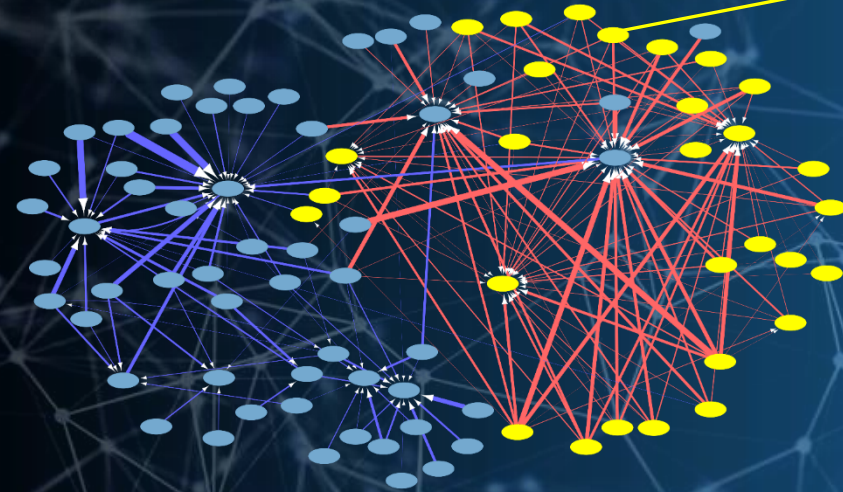
Case Study: Lung Cancer



- Assessment of KRAS/MEKK inhibition in Lung Cancer
- Top 200 drivers evaluated via analysis of 9 lung cancer datasets, generating 30 novel targets in KRAS (22 in MEKK)



Project led by Intellomx and supported by Innovate Grant. Subcontractors included: Medicines Discovery Catapult and cell-line robotics company Arctoris Ltd.



● Novel Intellomx discoveries
● Already known in literature

IO:03 KRAS inhibitor

IO:0301	Small molecule	wild type
IO:0302	Small molecule	wild type
IO:0303	Small molecule	wild type+G12C
IO:0304	Monoclonal	wild type+G12C

IO:04 KRAS/MEKK inhibitor

IO:0404	Small molecule	G12C
IO:0405	Small molecule	G12C

IO:02 MEKK inhibitor

IO:0202	Small molecule	wild type+G12C
---------	----------------	----------------

- 3 biological targets and 7 molecules prioritized for development and partnering

Case Study: IPF




- Question: are we able to identify causal factors/drivers of Idiopathic Pulmonary Fibrosis

- Publicly available transcriptomic data set analyzed GSE150910 - 103 IPF cases, 104 Non-IPF case
- Key pathway used as interrogation framework
Pathway: Cellular Senescence – 127 gene products
- Entire transcriptome analyzed for associations with selected pathway components

- Key features available for consideration in ranking:
 - Influence of a given gene product in the network
 - Influence of disease network on gene product
 - Degree of pathway connectivity
 - Stability of gene product across the population
 - Gene influence in healthy lung, liver and blood

- Work-in-progress ...



Gene Identifier	In Cellular Senescence list	Pathway Miner Connectivity	Connectivity Rank	Influence Rank IPF	Stability Score	OVERALL RANK	Protein Class
31	NO	20	114	49	44	1	apoptosis inhibitory protein
826	NO	18	190	32	81	2	metalloprotease
161	NO	21	76	158	85	3	oxidoreductase
139	NO	18	190	98	47	4	Golgi-localized complex
643	NO	21	76	177	99	5	Solute Carrier Protein
672	NO	17	228	116	12	6	C2H2 zinc finger transcription factor
666	NO	14	342	18	20	7	regulator of G protein signaling
376	NO	16	266	27	101	8	membrane traffic protein
293	NO	15	304	24	88	9	GTP-binding elongation factor
683	NO	18	190	131	102	10	RNA Binding Protein
279	NO	14	342	44	46	11	membrane traffic protein
601	NO	16	266	62	107	12	RING Finger Protein
599	NO	18	190	208	48	13	ubiquitin-protein ligase
470	NO	14	342	104	4	14	regulatory protein
201	NO	19	152	288	17	15	mannosyltransferase activator
4	NO	14	342	101	29	16	scaffold/adaptor protein
46	NO	20	114	341	23	17	transmembrane protein dislocase
412	NO	11	456	15	14	18	phosphatase
339	NO	11	456	26	5	19	scaffold/adaptor protein
734	NO	14	342	51	95	20	Kelch Like Family Member

Case Study: breast cancer



- Question: can we identify markers that predict proliferation in breast cancer?

JAMA Open
Association of Sperm-Associated Antigen 5 and Treatment Response in Patients With Estrogen Receptor-Positive Breast Cancer

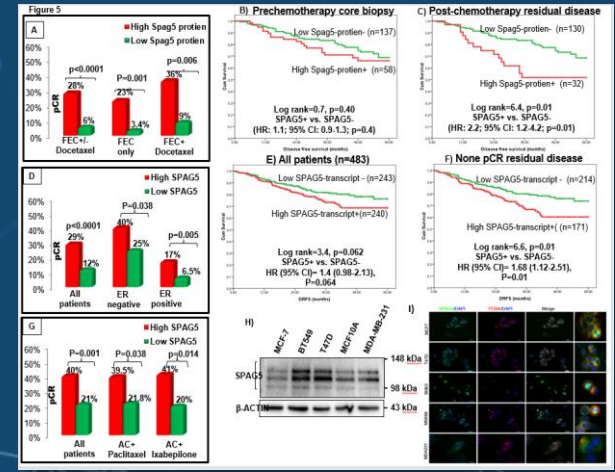
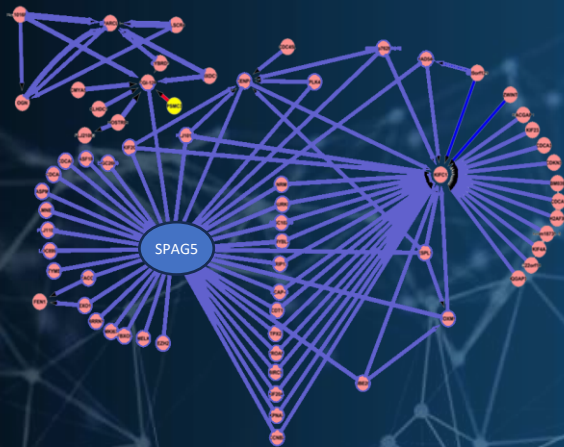
Abstract
 This is a prospective study that evaluates the optimal treatment of early-stage breast cancer by identifying the association between proliferation and treatment response.

OBJECTIVE
 To evaluate the association of sperm-associated antigen 5 (SPAG5) transcript and SPAG5 protein expression with treatment response in patients with estrogen receptor-positive breast cancer.

DESIGN, SETTING, AND PARTICIPANTS
 This retrospective cohort study included patients with estrogen receptor-positive breast cancer who received breast-conserving therapy with or without endocrine therapy as their first-line treatment. SPAG5 transcript and SPAG5 protein expression were measured in tumor tissue. The association of SPAG5 transcript and SPAG5 protein expression with pathologic complete response (PCR) was evaluated, as was the association of SPAG5 transcript and SPAG5 protein expression with treatment response. The association of these values for SPAG5 transcript and SPAG5 protein expression was assessed in a separate analysis.

RESULTS
 The study included 210 women aged 24 to 79 years from 2010 to 2016. SPAG5 transcript and SPAG5 protein expression were measured in tumor tissue. SPAG5 transcript expression was associated with pathologic complete response (PCR) (HR, 1.1; 95% CI, 1.0-1.2; p=0.004). SPAG5 protein expression was associated with treatment response (HR, 1.1; 95% CI, 1.0-1.2; p=0.004).

CONCLUSIONS
 SPAG5 transcript and SPAG5 protein expression were associated with pathologic complete response and treatment response in patients with estrogen receptor-positive breast cancer.



- Datasets: Nottingham, Uppsala, Metabric and TCGA breast cancer tissue expression array data. (3554 cases). 34 genes found consistently in the top 100 PANN ranked genes out of 50,000 genes across 5 questions across 3 data sets

Articles

SPAG5 as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis

Summary
 Background: Proliferation markers and profiles have been recommended for guiding the choice of systemic treatments for breast cancer. However, the best molecular marker or set to use has not yet been identified. We did this study to identify factors that drive proliferation and its associated features in breast cancer and assess their association with clinical outcomes and response to chemotherapy.

Methods
 We applied an artificial neural network-based integrative data mining approach to data from three cohorts of patients with breast cancer: the Nottingham dataset cohort (n=27), TCGA cohort (n=24), and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (n=76). We then identified the genes with the most effect on other genes in the resulting interaction map. Sperm-associated antigen 5 (SPAG5) featured prominently in this network.

Results
 We identified 34 genes consistently in the top 100 PANN ranked genes out of 50,000 genes across 5 questions across 3 data sets.

THE LANCET Oncology

Articles

SPAG5 as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis

Summary
 Background: Proliferation markers and profiles have been recommended for guiding the choice of systemic treatments for breast cancer. However, the best molecular marker or set to use has not yet been identified. We did this study to identify factors that drive proliferation and its associated features in breast cancer and assess their association with clinical outcomes and response to chemotherapy.

Methods
 We applied an artificial neural network-based integrative data mining approach to data from three cohorts of patients with breast cancer: the Nottingham dataset cohort (n=27), TCGA cohort (n=24), and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (n=76). We then identified the genes with the most effect on other genes in the resulting interaction map. Sperm-associated antigen 5 (SPAG5) featured prominently in this network.

- SPAG5 (Hub) validated as new driver of breast cancer

Case Study: TB Diagnostic Discovery



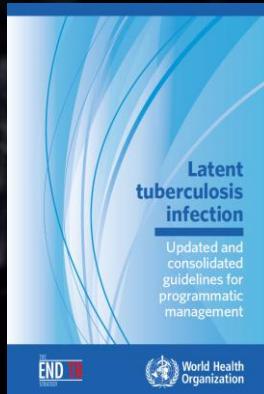
- **Question:** Can we create a diagnostic to differentiate between latent and active Tuberculosis?

“We believe this work will herald a new beginning for TB treatment and prevention.”



*Chairman Chen,
Wuhan Pulmonary
Hospital*

“There is no gold standard test for LTBI (Latent TB Infection).”



WHO 2024

Category	Test	Active TB	Latent TB ¹	BCG ²
Host response	TB-PRECISE	✓	✓	✓
Host response	MTB-HR	✓	✗	✓
Skin test	TST	✓	✗	✗
IGRA	QFT-Plus	✓	✗	✗
Live bacteria	Actiphage	✓	✗	✓

- 1 test provides sensitivity and specificity >85%
- 2 test results unaffected by patient BCG status

Note: Generalisation

Markers refined in the Intellomx Han Chinese population trial were originally identified in Caucasian and South African datasets using alternative data collection methods. This provides a high degree of confidence in the applicability of TB-PRECISE in the wider global population.



Validation:



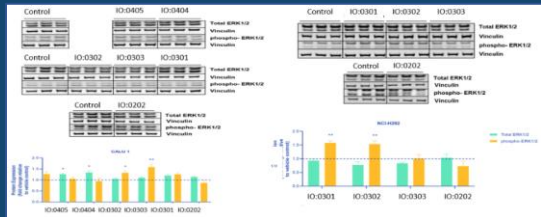
Intellomx output has been validated in multiple disease areas and clinical contexts. Examples:

Lung cancer

Breast cancer

Tuberculosis

Plant biology



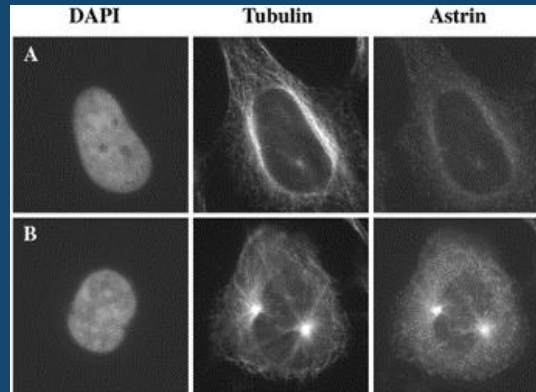
Validation:

WP 1.1 siRNA Screen: phenotypic outcome 24, 48, 72 and 96h
 WP 1.2 Pathway Analysis screen: siRNA transfection. Target depletion confirmed
 WP 2 Compound Screen: siRNA transfection. Target depletion results validated in Wild Type and Mutated cell lines

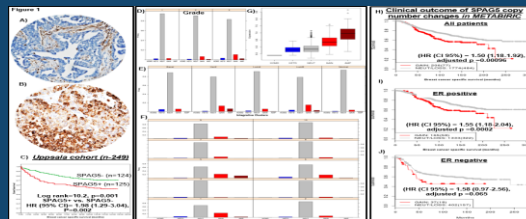
“Our system validated the targets and then tested drugs against those targets. Results were clear – the targets identified by Intelligent OMICS are indeed important, previously unknown targets in lung cancer and the novel drugs identified by the team modulate those targets, all as predicted by the AI.”



Martin Bittner
CEO



SPAG5 (Hub) validated in >15000 cases – Chemotherapy response.
 Cell line studies show functionality.
 $P(\text{false discovery}) < 1 \times 10^{-78}$



SPAG5: new driver of proliferation validated in over 15,000 cases. Patent Ref: US10775381B2

TB diagnostic tested in the Han population, showing high sensitivity and specificity. Work ongoing for a multi-site clinical trial in China.

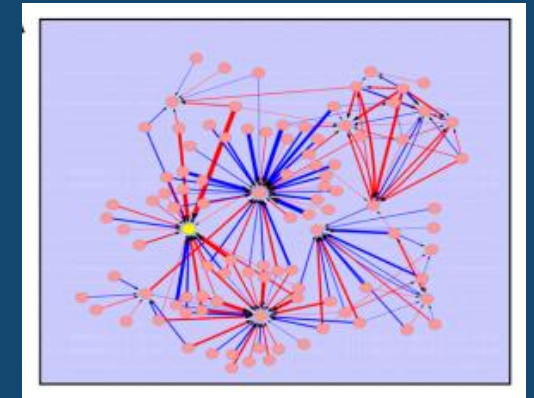
Trial results	Sensitivity	Specificity	AUC	Markers
Latent v Control	100%	87.9%	0.82	4
Active v Control	100%	100%	1.00	4
Active v Latent	100%	93.0%	0.91	5



Celebrating successful validation of our TB Diagnostic with the UK Dep Consul General in Wuhan, the the hospital Chairman and team



Two transcription factors were discovered using data mining and Network Inference. Development of transgenic fruit demonstrated on/off switching of ripening.



Intellomx I³ recap



I ³ -Distiller	<i>Identification of key genes</i>
I ³ -Driver	<i>Identification of key disease drivers</i>
I ³ -Miner	<i>Defining the disease pathway</i>

I³-Digital Twin *Anticipating tox liabilities*

I ³ -Pilot	<i>Identification of novel Small Molecules and Biologics</i>
I ³ -Precise	<i>Panel optimization for diagnostics</i>

The Intellomx I³ technology platform delivers validated targets, linked to *in silico* drug candidates.

Our algorithms deploy 5 layers of validation across multiple datasets, addressing data issues and dramatically reducing the probability of false discovery whilst our digital twins build comparator pathway models for healthy state in tissue of interest and in off-target tissues.

Results fill knowledge gaps based on evidence (whole transcriptome), quantifying the influence of each molecule in a pathway, identifying levels of dysregulation of each molecule and linking drug candidates with targets, ranked on multiple criteria.

The Intellomx I³ technology swarm-based approach¹ is not constrained by the availability of compute resource. Our swarms undertake multiple analyses, ranking results by concordance, with the option to drill into any single result or group of analyses to facilitate understanding and derive optimal drug targets.

1: see Appendix I for details of swarm technologies

Intellomx I³: Intuitive, Informed, Intelligence

Our Clients



Examples of corporate contracts:



Examples of Academic collaborations:



Key papers + illustrations:



Articles **Lancet oncology**
Supplementary data available on request

SPAG5 as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis

Tarek M A Abdel-Fattah¹, Devika Agrawal², Dong Wu Lu³, Rodin Ruzhik⁴, Oscar M Rivera⁵, Karen Liu⁶, Bing Xu⁶, Paul M Mosley⁶, Andrew H Green⁶, Alan Chacko⁶, Robert Evers⁶, Corina Galassi⁶, Ian D Ellis⁶, Graham R Babb⁷, Stephen Y T Chan⁸

Summary: Pathway analysis and protein data have been generated for defining the better treatment strategies.

REGULAR ARTICLES **blood advances**
A parsimonious 3-gene signature predicts clinical outcomes in an acute myeloid leukemia multicohort study

Sarah Wagner¹, Jayakumar Vadakekollu¹, Sarah K Tasian², Hadi Altamraz³, Martin Bornhäuser⁴, A Graham Pocock⁵, Graham R. Babb⁶, and Sergio Rottella⁷

Key Points

- Machine-learning approaches identified a parsimonious 3-gene expression signature that predicts risk in newly diagnosed AML.
- The 3-gene PI could be used to refine the accuracy of patient stratification and outcome prediction in routine clinical practice.

Acute myeloid leukemia (AML) is a genetically heterogeneous hematological malignancy with variable responses to chemotherapy. Although recurring cytogenetic abnormalities and gene mutations are important predictors of outcome, 50% to 70% of AMLs harbor normal or risk-indefinite karyotypes. Therefore, identifying more effective biomarkers predictive of treatment success and failure is essential for informing tailored therapeutic decisions. We applied an artificial neural network (ANN)-based machine learning approach to a publicly available data set for a discovery cohort of 593 adults with nonpromyelocytic AML. ANN analysis identified a parsimonious 3-gene expression signature comprising *CALCR1*, *CD109*, and *LSP1*, which was predictive of event-free survival (EFS) and overall survival (OS). We computed a prognostic index (PI) using normalized gene-expression levels and β -values from subsequently created Cox proportional hazards models, coupled with clinically established prognosticators. Our 3-gene PI separated the adult patients in each European LeukemiaNet cytogenetic risk category into subgroups with different survival probabilities and identified patients with very high-risk features, such as those with a high PI and either *FLT3* internal tandem duplication or mutated nucleophosmin 1. The PI remained significantly associated with poor EFS and OS after adjusting for established prognosticators, and its ability to stratify survival was validated in 3 independent adult cohorts ($n = 905$ subjects) and 1 cohort of childhood AML ($n = 145$ subjects). Further in silico analyses established that AML was the only tumor type among 39 distinct malignancies for which the concomitant upregulation of *CALCR1*, *CD109*, and *LSP1* predicted survival. Therefore, our ANN-derived 3-gene signature refines the accuracy of patient stratification and the potential to significantly improve outcome prediction.

Introduction

Acute myeloid leukemia (AML) is characterized by bone marrow (BM) and tissue infiltration by proliferative clonal abnormally differentiated cells of hematopoietic origin.¹ Prognosis is largely determined by cytogenetic abnormalities and AML-specific molecular lesions.² Although AML can be cured in 30% to 40% of adult patients aged <60 years with multiagent chemotherapy and often hematopoietic stem cell transplantation (HSCT), chemotherapy disease is common, and relapse represents a major cause of treatment failure.³ Investigation of new molecularly targeted agents for children and adults with high-risk AML remains a high priority.^{4,5}

Submitted 21 December 2018; accepted 13 March 2019; DOI 10.1182/bloodadvances.2018030724.
The full-text version of this article contains a data supplement.

© 2019 by The American Society of Hematology

frontiers in Oncology
ORIGINAL RESEARCH
Published: 26 August 2019
DOI: 10.3389/fonc.2019.01910

Frontiers in oncology
Comprehending Meningioma Signaling Cascades Using Multiscale Modeling

frontiers in Immunology
ORIGINAL RESEARCH
Published: 21 March 2020
DOI: 10.3389/fimm.2020.00490

Frontiers in biology
Development of a Bioinformatics Framework for Identification and Validation of Genomic Biomarkers

Plant physiology
Network Inference Analysis Identifies an *APRR2-Like* Gene Linked to Pigment Accumulation in Tomato and Pepper Fruits^{1|10|10A}

Yu Fan, Glyn Bradley², Kevin Pyke, Graham Ball, Chungui Lu, Rupert Fry, Alexandra Marshall¹, Subhalaj Jayasuta, Charles Baxter, Rik van Wijk, Laurie Boyden, Rebecca Cade, Natalie H. Chapman, Paul D. Fraser, Charlie Hodgman, and Graham B. Seymour³

Carotenoids represent some of the most important secondary metabolites in the human diet, and tomato (*Solanum lycopersicum*) is a rich source of these health-promoting compounds. In this work, a novel and trait-related regulator of pigment accumulation in tomato has been identified by artificial neural network inference analysis and is function validated in transgenic plants. A tomato fruit gene regulatory network was generated using artificial neural network inference analysis and transcription factor gene expression profiles derived from fruits sampled at various points during development and ripening. One of the transcription factor gene expression profiles with a sequence related to an Arabidopsis (*Arabidopsis thaliana*) *ARABIDOPSIS FUS3/REPRESSOR REGULATORY-LIKE* gene (*APRR2-Like*) was up-regulated at the breaker stage in wild-type tomato fruits and, when overexpressed in transgenic lines, increased plastid number, area, and pigment content, enhancing the levels of chlorophyll in immature stripe fruits and carotenoids in red ripe fruits. Analysis of the transcriptome of transgenic lines overexpressing the tomato *APRR2-Like* gene revealed up-regulation of several ripening-related genes in the overexpressing lines, including a link between the expression of this tomato gene and the ripening process. A putative ortholog of the tomato *APRR2-Like* gene in sweet pepper (*Capiscum annuum*) was associated with pigment accumulation in fruit tissues. We conclude that the function of this gene is conserved across taxa and that it encodes a protein that has an important role in ripening.

Tomato (*Solanum lycopersicum*) is a climacteric fruit whose ripening is initiated and coordinated by ethylene (Alexander and Grierson, 2002). It is the model system

for studying ripening in fleshy fruits because of the extant genetic and molecular resources that are available, including well-characterized mapping populations (Leppman et al., 2007), numerous single-gene mutants, routine transformation, and a fully annotated genome sequence (Tomato Genome Consortium, 2012). The repertoire of well-characterized mutations in tomato has permitted the identification of genes that encode proteins that govern the ripening process. These have included *Non-ripening 1* (*Nr1*), *Tripartite inhibitor 1* (*Tri1*), *non-ripening 2* (*Nr2*), and *Colorless nonripening 2* (*Cnr2*). Mutations at these loci can completely abolish normal ripening (Lanahan et al., 1994; Vrebalov et al., 2002; Manning et al., 2006). The *Nr*, *Ri*, *Cnr*, and *Ncr* gene products, along with those from tomato *Hd-Zip* (*homotetraploid 1*) (*Hd1*), *Tomato AGAMOUS-LIKE1* (*TAGL1*), *APETALA2* (*Ap2*), Liu et al., 2008; Ikkan et al., 2009; Vrebalov et al., 2009; Chung et al., 2010; Karlova et al., 2011), and others govern the onset and progression of the ripening. Despite a growing understanding of this high-level regulatory network, the links to hormonal cues, plastid signals, and downstream effectors mediating alterations in color, texture, and flavor are still poorly understood.

105 Plant Physiology, March 2019, Vol. 141, pp. 1476-1483, www.plantphysiol.org © 2019 American Society of Plant Biologists. All Rights Reserved. Downloaded from on April 3, 2020. Published by www.plantphysiol.org Copyright © 2019 American Society of Plant Biologists. All rights reserved.

frontiers in Immunology
ORIGINAL RESEARCH
Published: 21 March 2020
DOI: 10.3389/fimm.2020.00490

Frontiers in biology
Development of a Bioinformatics Framework for Identification and Validation of Genomic Biomarkers

JAMA Network Open
Original Investigation | Oncology
Association of Sperm-Associated Antigen 5 and Treatment Response in Patients With Estrogen Receptor-Positive Breast Cancer

Tarek M. A. Abdel-Fattah, PhD, Graham R. Babb, PhD, Paul U. Thangavelu, PhD, Lynne E. Reid, PhD, Amy E. McCart Reed, PhD, Judith M. Searles, PhD, Pascal H. G. Doff, PhD, Peter J. Simpson, PhD, Sarah L. Lahari, MD, Louise Forgue, PhD, Balaji Gylfry, PhD, Paul M. Mosley, BSc (Hons), Andrew H. Green, PhD, Alan Chacko, MD, Ian D. Ellis, MD, Stephen Y. T. Chan, MD

Abstract

IMPORTANCE: There is no proven test that can guide the optimal treatment, either endocrine therapy or chemotherapy, for estrogen receptor-positive breast cancer.

OBJECTIVE: To investigate the associations of sperm-associated antigen 5 (SPAG5) transcript and SPAG5 protein expressions with treatment response in systemic therapy for estrogen receptor-positive breast cancer.

DESIGN, SETTINGS, AND PARTICIPANTS: This retrospective cohort study included patients with estrogen receptor-positive breast cancer who received 5 years of adjuvant endocrine therapy with or without neoadjuvant anthracycline-based combination chemotherapy (NACT) derived from 11 cohorts from December 1, 1986, to November 28, 2019. The associations of SPAG5 transcript and SPAG5 protein expression with pathological complete response to NACT were evaluated, as was the association of SPAG5 mRNA expression with response to neoadjuvant endocrine therapy. The association of SPAG5 mRNA expression with response to neoadjuvant endocrine therapy and the association of SPAG5 mRNA expression with response to neoadjuvant endocrine therapy were analyzed. Data were analyzed from September 9, 2015, to November 28, 2019.

MAIN OUTCOMES AND MEASURES: The primary outcomes were breast cancer-specific survival, distal relapse-free survival, pathological complete response, and clinical response. Outcomes were examined using Kaplan-Meier, multivariable logistic, and Cox regression models.

RESULTS: This study included 12 720 women aged 24 to 78 years (mean [SD] age, 58.46 [14.52] years) with estrogen receptor-positive breast cancer, including 1073 women with SPAG5 transcript expression and 361 women with SPAG5 protein expression of locally advanced disease stage IIA through IIC. Women with SPAG5 transcript and SPAG5 protein expressions achieved higher pathological complete response compared with those without SPAG5 transcript or SPAG5 protein expressions (transcript: odds ratio, 2.45 [95% CI, 1.73-3.51]; $P < .001$; protein: odds ratio, 7.32 [95% CI, 3.23-16.22]; $P < .001$). Adding adjuvant anthracycline chemotherapy to adjuvant endocrine therapy for SPAG5 mRNA expression in estrogen receptor-positive breast cancer was associated with prolonged 5-year distal relapse-free survival in patients without lymph node involvement (hazard ratio, 0.34 [95% CI, 0.14-0.87]; $P = .033$) and patients with lymph node involvement (hazard ratio, 0.35 [95% CI, 0.18-0.68]; $P = .002$) compared with receiving 5-year endocrine therapy alone. Mean OS/SPAG5 transcript was found to be downregulated after 2 weeks of neoadjuvant endocrine therapy compared with pretreatment levels in 68 of 92 patients (74%) (0.23 [0.18] vs 0.34 [0.24]; $P < .001$).

Open Access. This is an open access article distributed under the terms of the CC BY License. JAMA Network Open. 2020;3(7):e200486. doi:10.1001/jamanetworkopen.2020.0486

obn Issue 5 Autumn 2021
CONTACT US: Please contact the OBN team: team@obn.org.uk if you would like to contribute to the next CONNECT or place an advert.
MEMBERSHIP DIRECTORY: An easy reference tool to see all of the OBN current Members and their offerings.

OBN supporting and bringing together the UK's life sciences companies, corporate partners and investors

CONNECT
DDI & drug

Intelligent OMICS, UKRI, Innovate UK, CATAPULT Medicines Discovery, ARCTORIS

Conquering KRAS!
The AI drug discovery revolution that is here to stay

Intelligent OMICS is delighted to announce successful completion of its recent Innovate UK grant program, in collaboration with the Medicines Discovery Catapult (MDC) and Arctoris Limited.

5 key conclusions:

- Application in oncology identifies novel KRAS-inhibiting drugs for lung cancer, one of which is even effective regardless of KRAS mutation.
- AI-driven drug discovery is at last proven, and ready for application across multiple diseases.
- The AI approach discovers new drugs that it focuses on the drivers or causes of the disease state, rather than focussing on the symptoms or effects of a disease state.
- The evidence-based analysis produces original results, without reliance on prior hypotheses or literature, allowing creation and control of new intellectual Property.
- AI discovery methods can achieve greater than 90% reduction in carbon footprint compared to traditional high-throughput screening.

The project, led by Intelligent OMICS and funded by Innovate UK, sought to demonstrate the carbon efficiency of an AI-based drug discovery program compared to traditional pharma methods. The case study used in the project was assessment of Non-Small Cell Lung Cancer – thought to account for over 80% of all lung cancer cases.

Method

The team analyzed nine lung cancer datasets from the Intellimaps Curated Data Library, using the Intelligent OMICS platform. The datasets include human transcriptomic data plus confirmation of a disease's healthy diagnosis for lung cancer for approximately 2,000 patients. Proprietary AI was used to model the underlying systems biology – first creating a list of the most important genes defining the disease's healthy diagnosis, then modelling the interaction of those genes in a disease pathway map based on the evidence in the data.

"The real benefit of our technology is evident when we compare our results with what is known in the literature," says Intellimaps CEO Dr Simon Fawcett. "We can immediately spot errors and omissions in pathway maps documented in KEGG, for example, and because our analysis only focusses on the most influential drivers in each pathway we know that any such differences are genuinely important. For our lung cancer work, focussing on EGFR and KRAS, the comparison led us to 8 really exciting new lung cancer targets."

The next step of the process was to validate in silico targets in the wet lab, to link validated targets to possible drugs and then to test the impact of those novel drugs on cancer cells.




Subcontractor Arctoris, with its world leading fully automated drug discovery platform and robotic cell line system, provided rapid validation of the targets using knock down analysis on KRAS G12C mutant and KRAS wild type cell lines. Data from Arctoris proved the validity of the targets on

Downloaded From: https://jamanetwork.com/ on 07/07/2020

Press service


Differentiation vs Large Language Models and other AI approaches



<p>Traditional Pharma Model</p> 	<p>Slow and expensive Lab-based Constrained by hypotheses</p>
<p>1st Generation AI</p> 	<p>Based on LLMs, Deep Learning or Binary Classifiers High computational requirement - inefficient Prone to false discovery and interpretation. Difficult to explain Over-reliant on known biology</p>
<p>2nd Generation AI</p> 	<p>Swarm-Based Neural Networks – Optimised parallel computing Extensive cross-validation through concordance Explainable – Knowledge Graph easily interpretable Evidence-based on mathematics, not language Reveals the hidden biology of disease Multi-parameter probability ranking – optimised drug targets</p>

Differentiation vs Large Language Models and other AI approaches



	Benefit	 Intellomx	InSilico	Exscientia	Owkin	In Vitro	Precision Life
Evidence based, data driven approach to biomarker discovery	<ul style="list-style-type: none"> • Avoids bias of LLMs • Reveals novel biology • Eliminates false discovery 	✓	✗	✗	✓	✓	✓
Network Inference modelling determines molecular causality	<ul style="list-style-type: none"> • Quantifies pathway • Reveals drivers • Determines dysregulation • Facilitates digital twin 	✓	✗	✗	✗	✗	✗
Reliant on human disease tissue	<ul style="list-style-type: none"> • Clinically relevant • Eliminates bias/downstream failure of cell/animal models 	✓	✗	✗	✓	✗	✓
Target biomarker discovery matrix	<ul style="list-style-type: none"> • Systematically evaluates whole transcriptome • Rapidly determines actionable targets (10 parameters) 	✓	✗	✗	✗	✗	✗

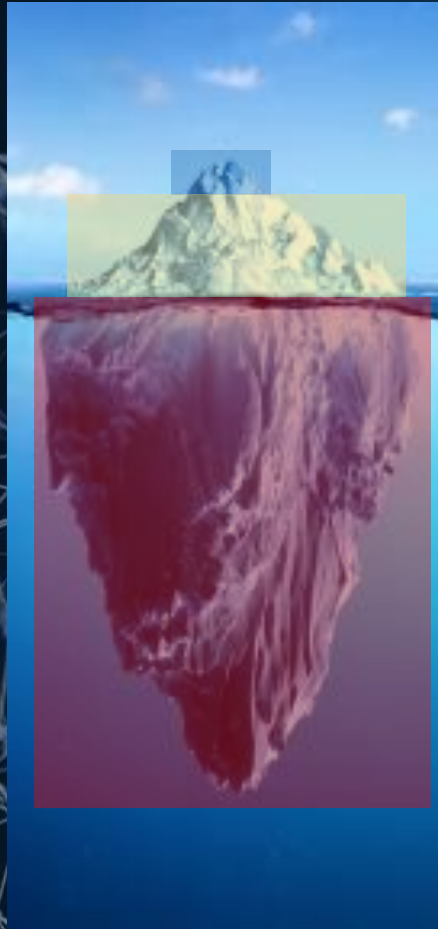
Differentiation vs Large Language Models and other AI approaches



Discovered and in the literature

Discoverable by LLMs from the literature

de novo discoveries from Molecular data



We DON'T rely on flawed reductionist knowledge graphs or literature mining to identify the key drivers of each disease.

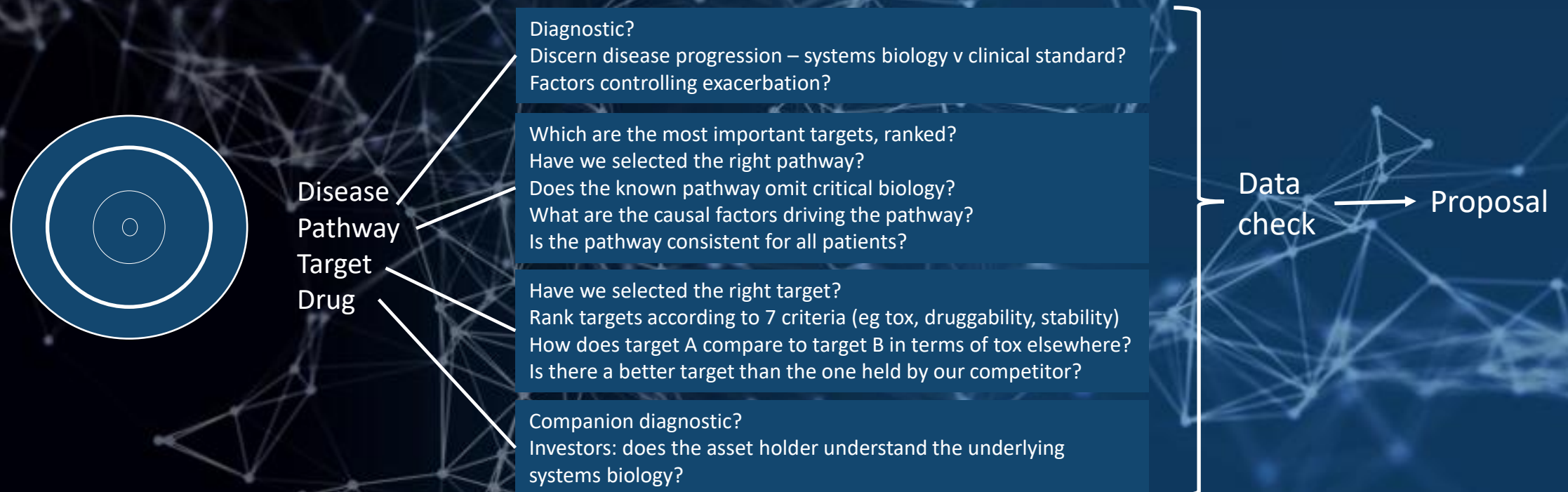
Instead, we use de novo, evidence-based modelling that discovers the underlying systems biology of each disease by direct analysis and then exploits that knowledge to create a stream of new drugs.

What happens next?



From first contact to project initiation typically takes 6 weeks. At first meeting the Intellomx team explains the utility and validation of the I³ tool box, reveals prior work in the disease area of interest and reviews what progress has already been made by the partner company. The team then helps clarify the clinical questions that need to be answered (see common queries at each research stage below).

Following the meeting, the Intellomx team reviews possible data sources and provides a formal proposal for review.



Contact



“To complete a long journey, it really helps if one sets off in the right direction. In our sector that basically means **selecting the right target**.

Our AI reveals the underlying biology of disease in order to identify optimum drug targets. *In silico* docking identifies drug candidates, and engagement with the full power of downstream AI enables us to de-risk drug development.

This is the decade in which the IP for a stream of new drugs will be secured, reversing the troubling trend in drug discovery globally. **Our technologies write the papers that LLMs will eventually read.”**

Professor Graham Ball, Founder, Intellomx



Contact:

Dr. Simon Haworth	simon.haworth@intellomx.com	+1 236 989 9555
Prof. Graham Ball	graham.ball@intellomx.com	+44 7817 189835
Dr. Bill Mason	bill.mason@intellomx.com	+44 7785 950134





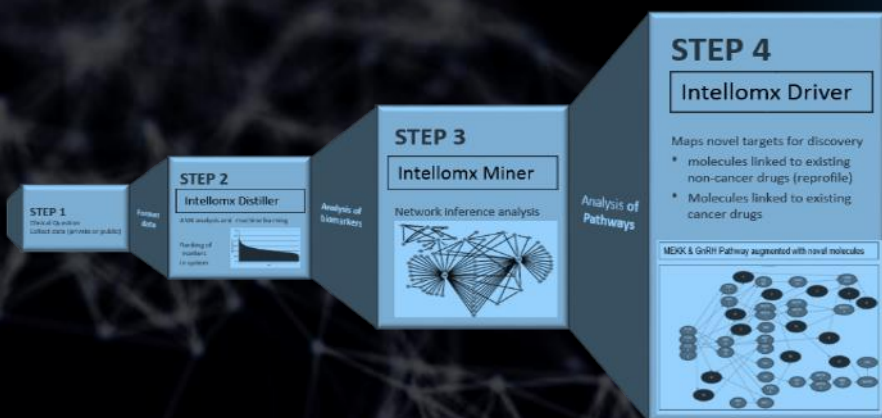
drug discovery AI

APPENDIX – Swarm Based AI Approaches

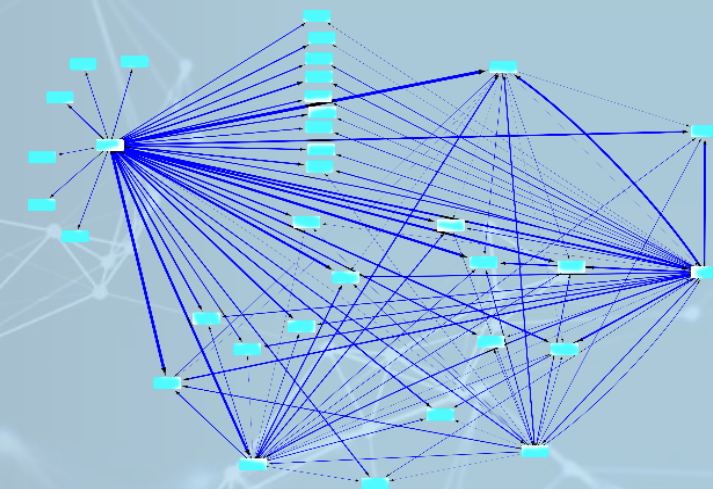
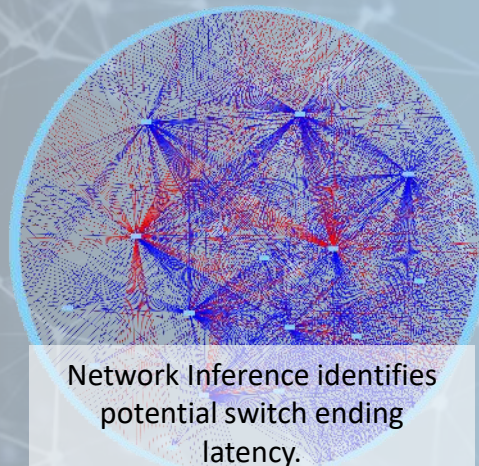


- Trains large numbers of smaller less deep Neural Network Models
- Each model takes a different view of the data or addresses a different part of the problem
- Data presentation allows the whole problem to be represented with overlaps
- Early stopping and regularisation and Monte Carlo cross-validation built in.
- >1000 models run at the same time on different compute units of a GPU
- All of the transcriptome is considered in parts, starting from a single gene product

Case study: Schizophrenia

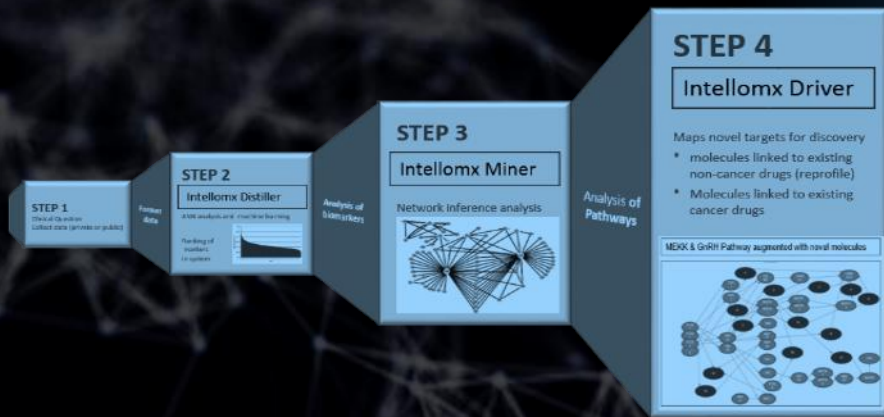


- Question: can we identify novel, key markers for schizophrenia and link these to potential drug compounds?



Primary analysis identified a small number of critical hubs that drive disease state, representing drug targets. NCEs and drugs for reprofiling have been identified that are linked to these targets.

Case study: COPD



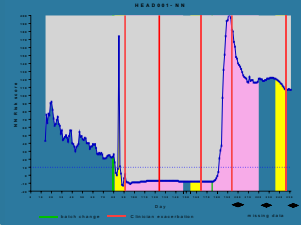
Question: can we improve the Mologic COPD panel for application on hand-held device in a clinical trial?

Intellomx POSitive service:

Mologic's COPD panel was highly effective but impractical to use due to high numbers of biomarkers. The Intellomx POSitive service enabled the panel to be reduced to 12 biomarkers.

Intellomx built a decision support model and code enabling the customer to deploy the companion diagnostic on mobile phone apps.

Original panel: 500 biomarkers



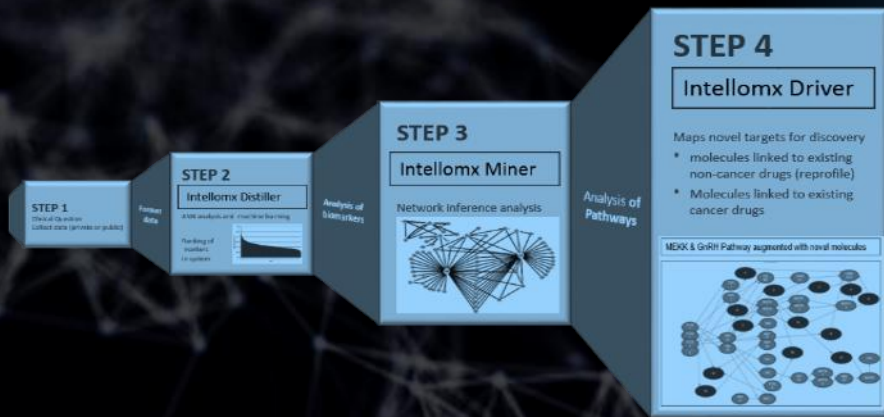
Refined panel:

12
12 features of 5 biomarkers



Conclusion: Existing panel optimised and software provided for application on hand-held devices/mobile phone application

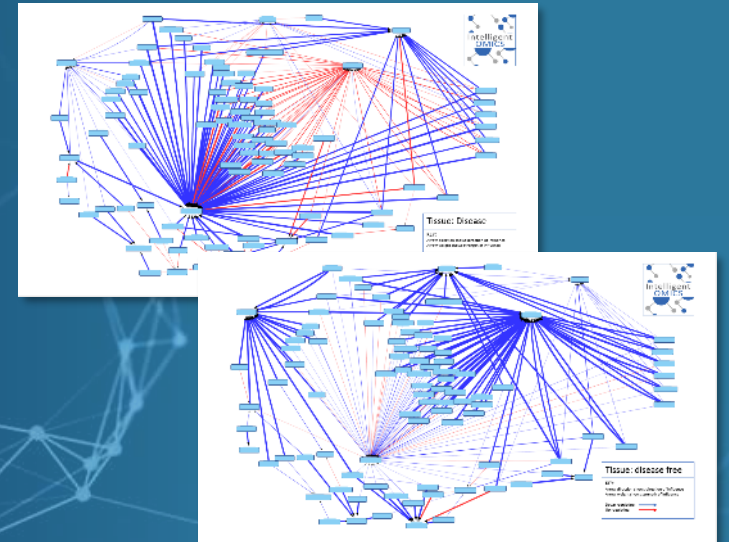
Case study: Auto-immune Disease



Question: Can we develop new drugs for autoimmune disease?

Intellox completed Primary Analysis for Autoimmune disease based on both client data and publicly available data, identifying 156 key genes mapped in the disease pathway.

The pharma co provided details of two targets:
Target 1 was validated and linked to the disease pathway
Target 2 was shown to be of negligible relevance in the disease



Analysis of a complex system to explain phenotype and stratification