uniquely enabling value based data navigation

eagle genomics

e[discover]

Best of Show Award at Bio-IT World Conference & Expo 2016

BEST OF SHOW Bio-IT World WINNER

e(discover) objectively and statistically measures the value of data assets as defined by their usefulness and relevance. The approach is based on decision theory and exploits quantitative and probabilistic techniques to enable a unique conversational/question driven approach, allowing scientists to explore scientific value across diverse data sets as never previously possible.

HOW IT WORKS

e(discover) is a first-generation tool capable of objectively assessing data value (as opposed to "quality" or monetary value) of scientific datasets using Decision Theory techniques. A hierarchical model is used to match datasets (value components) to the question posed by the data scientist. Once modelling is complete, **e**(discover) allows the user to apply value to the catalogue, discover relationships between components contributing to the value, and to further refine the model, never previously possible.

BUSINESS BENEFITS

Guides information management strategies
By providing users with prioritized data based on scientific value, **e**(discover) (working in conjunction with **e**(catalog)) improves data selection, encourages data reuse and informs experimental programs.

Breaks the loop of data-analysis paralysis
eaglediscover addresses the challenge where too much time is spent on data wrangling to the detriment of scientific interpretation thus breaking the loop of data-paralysis.

Capture tacit knowledge
The ability to capture and use tacit knowledge in a scientific or business decision context increases ROI on acquired knowledge across multiple stages of product development.

Closes the Design of Experiments Loop
Valuation of data within a internally coherent framework is bound to support improved experimental design. Low value (as opposed to only low quality) experiments can be easily flagged and drive leaner more informative experiments.

MODULE FEATURES

Unique question-driven approach » allows scientists to explore datasets using validated scientific queries.

Objective guide to data curation and enhancement

Data selection by prioritization (scoring) » not simply down-selection (filtering).

Hierarchical search & discovery visualization

Enterprise integration with proprietary or public data sources

Translation of tacit knowledge to a business experimental process » subject to PDSA learning cycle.

Conversational environment (human-machine & human-human) » for refining data value model.

A USE CASE – INTERNATIONAL CANCER GENOME CONSORTIUM DATA

The **e**(discover) has demonstrated novel insight on the world's largest public cancer patient. ICGC is a public resource used globally by scientists to understand and find cures for cancer. Eagle Genomics' analysis enabled a ranking of the most valuable projects according to their scientific value.

The **e**(discover) product objectively measures the value of data assets as defined by their usefulness and relevance. The approach exploits quantitative and probabilistic techniques and conversation theory to enable a unique question-driven approach, allowing scientists to explore scientific value across diverse data sets as never previously possible.

Results

Below are some examples of the reporting from the study (fig. 2), along with a snapshot of the data valuation, it provides a scoring of all records. We show the top 20 in the list (fig. 3).

Fig. 1

The workflow for the ICGC study

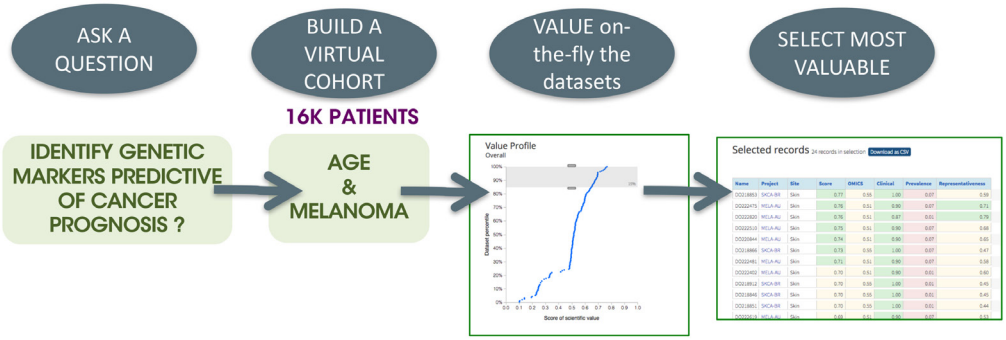


Fig. 2

Omics data

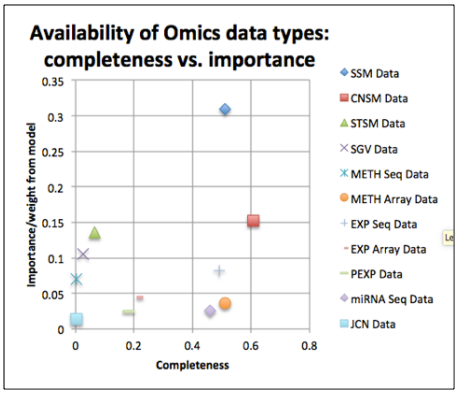


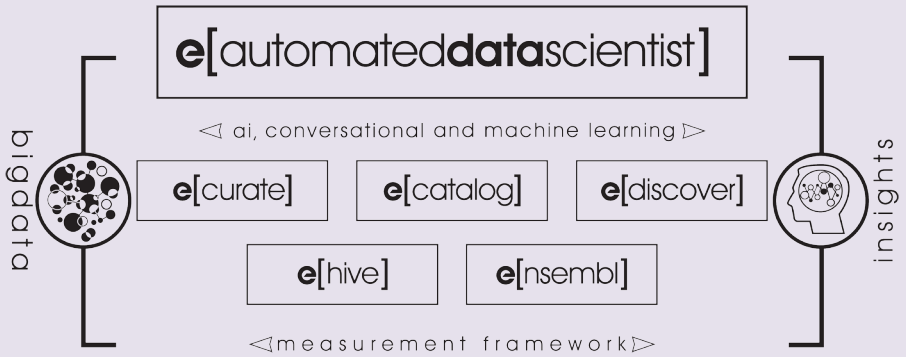
Fig. 3

Snapshot of data valuation

Name	Project	Site	Score	OMICS	Clinical	Prevalence	Representativeness
DO4671	BRCA-US	Breast	1.00	0.81	0.58	1.00	0.53
DO2966	BRCA-US	Breast	1.00	0.81	0.58	1.00	0.52
DO2874	BRCA-US	Breast	1.00	0.78	0.58	1.00	0.55
DO5654	BRCA-US	Breast	1.00	0.73	0.58	1.00	0.60
DO3530	BRCA-US	Breast	0.99	0.78	0.51	1.00	0.60
DO5936	BRCA-US	Breast	0.99	0.81	0.58	1.00	0.50
DO3988	BRCA-US	Breast	0.99	0.78	0.58	1.00	0.52
DO2860	BRCA-US	Breast	0.99	0.81	0.58	1.00	0.49
DO5081	BRCA-US	Breast	0.98	0.78	0.58	1.00	0.50
DO1331	BRCA-US	Breast	0.98	0.81	0.51	1.00	0.54
DO4080	BRCA-US	Breast	0.97	0.78	0.58	1.00	0.48
DO4866	BRCA-US	Breast	0.97	0.81	0.51	1.00	0.52
DO4713	BRCA-US	Breast	0.97	0.78	0.58	1.00	0.48
DO2761	BRCA-US	Breast	0.97	0.73	0.58	1.00	0.53
DO5060	BRCA-US	Breast	0.97	0.81	0.51	1.00	0.50
DO1396	BRCA-US	Breast	0.97	0.78	0.58	1.00	0.46
DO1250	BRCA-US	Breast	0.96	0.78	0.51	1.00	0.52
DO5808	BRCA-US	Breast	0.96	0.81	0.51	1.00	0.49
DO6056	BRCA-US	Breast	0.96	0.73	0.58	1.00	0.50
DO44248	BRCA-US	Breast	0.96	0.73	0.58	1.00	0.50

OUR SMART DATA MANAGEMENT PLATFORM

e(discover) is part of our proprietary software suite that bridges the entire process from the medical data through to insight.



"This 'conversation' between the scientist and the data sets is the next wave of innovation that we need deployed to our R&D teams, so we can quickly and systematically find and validate new compounds in this Precision Medicine era. Eagle is very much focused on solving this problem and we fully support their product plans."

Mathew Woodwark, Director of Bioinformatics, MedImmune

Tel: +44 (0) 1223 654481

Email: info@eaglegenomics.com

The Biodata Innovation Centre, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1DR, UK.