The Advanced Data Analysis Centre (ADAC)

- Background
- Aims and Objectives
- Areas of expertise

Current projects
- NPI+ (Breast Cancer)
- WildTech
- Alzheimer’s disease
Background

- Increasingly sophisticated methods of data analysis are becoming mainstream
  - Analysing and presenting data in research papers
  - Analysis work-packages in research projects

- We (IMA / CS) are experts in data analysis
  - Make this expertise available as a service across the University to support others’ research
ADAC

- A joint initiative between Vet School and CompSci
- Provide data analysis service across University
- Support high-class research
  - Analysis and interpretation of complex data
  - Development of research strategy
  - Grant writing

In Computer Science
- Jon Garibaldi
- Uwe Aickelin
- Daniel Soria
- Grazziela Figueredo
Specific Objectives

- Analyse data
  - Difficult data (beyond the normal comfort-zone of the researchers possessing the data) providing services as required by the clients

- Provide data analysis assistance / training
  - ADAC will provide assistance and training for clients

- Assist in writing funding applications and research outputs
  - Writing of research articles, for increased impact of publications
  - Writing of work packages involving data analysis for funded projects

- Provide advice to clients on
  - Data collection methods
  - Statistical or other data analysis methods
  - Presentational methods to be used to increase impact
  - Ethical and legal issues surrounding data collection, storage and analysis
20th Century Death

Ischemic heart disease 540m
Stroke 410m
Other cardiovascular 178m
Non-communicable diseases 1,970m

Humanity 980m

War 137m
WWII 60m
WWI 37m

Accidents 298m
Deaths from accidents 115m

Drugs 115m
Illegal drugs 65m
Tobacco 30m

Stomach 46m
Liver 46m
Air pollution 116m
Mortality from pollution 30m

Ideology 342m
Communism 94m
China 65m
North Korea 28m

Ideology 342m
Genocide 9m
Rwanda 500,000
Cameroon 100,000

Homicide 68m
Gay marriage 9m

Mortality from violence 177m
Murder 177m
Suicide 59m

Epidemics 1,680m

Information is Beautiful
Areas of expertise

- Artificial Intelligence for decision support
- Machine Learning methodologies
- Data Management
- Simulation
- Bioinformatics
- System Biology
ADAC Impact

ADAC

publications

impact

funding

Advanced Data Analysis Centre

28/05/2013
Advanced Data Analysis Centre

What is ADAC?
The Advanced Data Analysis Centre (ADAC) is funded to support research by the analysis and interpretation of complex data.

Why is this important?
Complex data underpins current research, while innovative analysis can prompt new discoveries.

Excellent research can often be stalled due to a lack of expertise in conducting data analysis. ADAC supports high-class research, providing analysts to help with the development of a research strategy, grant writing, and interpretation of complex data.

Case Studies
- View our current case studies from the Expertise page

Quick links
- The University of Nottingham
- School of Veterinary Medicine and Science
- School of Biosciences
- School of Computer Science

Contact us
- General queries
- Bioinformatics queries
- Informatics queries

Advanced Data Analysis Centre
ADAC is a joint initiative between the School of Computer Science and the School of Veterinary Medicine and Science.
Please see our Contact Us page for telephone and email contact points.
On-going projects

- **NPI+ (breast cancer)**
  - Refine biological classes
  - Develop new NPI score for personalised treatment

- **WildTech**
  - EU-FP7 funded project, Vet School & partners across Europe
  - Build and maintain an online database system

- **Alzheimer disease**
  - DPFS-MRC project, School of Biology
  - Classification of patients using small panel of biomarkers
Aim

To establish model-based techniques to classify patients into given classes

Methodology

Fuzzy logic used to provide indication of the likelihood of membership into classes

Output

Algorithm required to place new patients into the treatment classes
Consensus Clustering

Breast Cancer

ER+ Luminal CKs+  ER- Basal CKs-  ER- HER2+

Luminal  Basal  HER2 Class 6 (7.2%)

PgR+  PgR-  p53+  p53-

HER3+  HER3-  HER3+  HER3-  HER4+  HER4-

Luminal A Class 1 (18.8%)  Luminal N Class 2 (14.2%)  Luminal B Class 3 (7.4%)

Soria et al., *Computers in Biology and Medicine*, 2010
Crisp sets

Integers lower than 20
Crisp sets? 

\[ \mu \]

0.00 0.25 0.50 0.75 1.00

0 20 40 60 80 100

Young
Fuzzy sets

μ

0.00
0.25
0.50
0.75
1.00

0
20
40
60
80
100

Young

28/05/2013
Fuzzy sets

μ

0.00 0.25 0.50 0.75 1.00

0 20 40 60 80 100

Young
Middle-Aged
Old

28/05/2013
Fuzzy biomarkers
Algorithm specifications

- Ten variables (biological markers)
  - ER, PgR
  - CK 7/8, CK 5/6
  - EGFR, HER2, HER3, HER4
  - p53, MUC1

- Different membership functions

- Cut-off points

- Output
  - List of possibilities (membership to each class)
  - NC < 5%
  - 12.5% < HER2+ < 15%
  - Linguistic rules table
Fuzzy logic provides a complete mathematical framework for combining terms of different variables together in the form of fuzzy rules:

- IF ER IS High AND PgR IS High …
  THEN Class IS Class_1
- IF ER IS Low AND HER2 IS High …
  THEN Class IS Class_6B
- …
T-norms and t-conorms

- Min – max

\[ T_{\text{min}}(a, b) = \min(a, b) \]
\[ \perp_{\text{max}}(a, b) = \max(a, b) \]

- Product – probabilistic sum

\[ T_{\text{prod}}(a, b) = a \cdot b \]
\[ \perp_{\text{sum}}(a, b) = a + b - ab \]
How the algorithm works (1)

- Using variables membership functions, data are changed into ‘low-high’

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient</td>
<td>er</td>
<td>pgr</td>
<td>ck78</td>
<td>ck56</td>
<td>egfr</td>
<td>cerbb2</td>
<td>cerbb3</td>
<td>cerbb4</td>
<td>p53</td>
<td>muc1</td>
</tr>
<tr>
<td>1605</td>
<td>60</td>
<td>0</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>170</td>
<td>100</td>
<td>150</td>
<td>180</td>
</tr>
<tr>
<td>1625</td>
<td>280</td>
<td>0</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>220</td>
<td>125</td>
<td>0</td>
<td>235</td>
</tr>
<tr>
<td>1628</td>
<td>215</td>
<td>110</td>
<td>156</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>50</td>
<td>200</td>
<td>0</td>
<td>170</td>
</tr>
<tr>
<td>1636</td>
<td>240</td>
<td>300</td>
<td>266</td>
<td>45</td>
<td>50</td>
<td>90</td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>290</td>
</tr>
<tr>
<td>1637</td>
<td>95</td>
<td>0</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>150</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>1638</td>
<td>210</td>
<td>75</td>
<td>233</td>
<td>53</td>
<td>30</td>
<td>65</td>
<td>90</td>
<td>50</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>1641</td>
<td>0</td>
<td>0</td>
<td>213</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>150</td>
<td>100</td>
<td>205</td>
</tr>
<tr>
<td>1644</td>
<td>100</td>
<td>190</td>
<td>266</td>
<td>80</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>220</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient</td>
<td>er</td>
<td>pgr</td>
<td>ck78</td>
<td>ck56</td>
<td>egfr</td>
<td>cerbb2</td>
<td>cerbb3</td>
<td>cerbb4</td>
<td>p53</td>
<td>muc1</td>
<td>cerbb2</td>
<td>cerbb2</td>
</tr>
<tr>
<td>1605</td>
<td>0.78245</td>
<td>0.21755</td>
<td>0.879107</td>
<td>0.120893</td>
<td>0.916827</td>
<td>0.083173</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.000335</td>
<td>0.999665</td>
</tr>
<tr>
<td>1625</td>
<td>0.003141</td>
<td>0.996859</td>
<td>0.879107</td>
<td>0.120893</td>
<td>0.310026</td>
<td>0.689974</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.832018</td>
<td>0.167982</td>
</tr>
<tr>
<td>1628</td>
<td>0.024602</td>
<td>0.975398</td>
<td>0.177118</td>
<td>0.822882</td>
<td>0.978289</td>
<td>0.021711</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.420676</td>
<td>0.579324</td>
</tr>
<tr>
<td>1636</td>
<td>0.011206</td>
<td>0.988794</td>
<td>0.000492</td>
<td>0.999508</td>
<td>0.571506</td>
<td>0.428494</td>
<td>0.310026</td>
<td>0.689974</td>
<td>0.276878</td>
<td>0.723122</td>
<td>0.21755</td>
<td>0.78245</td>
</tr>
<tr>
<td>1637</td>
<td>0.539915</td>
<td>0.460085</td>
<td>0.879107</td>
<td>0.120893</td>
<td>0.310026</td>
<td>0.689974</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.000335</td>
<td>0.999665</td>
</tr>
<tr>
<td>1638</td>
<td>0.028748</td>
<td>0.971252</td>
<td>0.397474</td>
<td>0.602526</td>
<td>0.793147</td>
<td>0.206853</td>
<td>0.258075</td>
<td>0.741925</td>
<td>0.420676</td>
<td>0.579324</td>
<td>0.382252</td>
<td>0.617748</td>
</tr>
<tr>
<td>1641</td>
<td>0.960834</td>
<td>0.039166</td>
<td>0.879107</td>
<td>0.120893</td>
<td>0.879107</td>
<td>0.120893</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.000335</td>
<td>0.999665</td>
</tr>
<tr>
<td>1644</td>
<td>0.5</td>
<td>0.5</td>
<td>0.016367</td>
<td>0.983633</td>
<td>0.571506</td>
<td>0.428494</td>
<td>0.127862</td>
<td>0.872138</td>
<td>0.654753</td>
<td>0.345247</td>
<td>0.000335</td>
<td>0.999665</td>
</tr>
</tbody>
</table>
How the algorithm works (2)

- Create rules from training data using t-norm and t-conorm
- Number of rules according to the number of possible classification outcomes
- Apply these rules to testing data
  - Class membership algorithm
  - Class assignment algorithm
Class membership algorithm

- Using again t-norm and t-conorm, values from fuzzy rules and from testing data are compared.
- List of memberships, each representing the possibility of the patient to belong to that particular class.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patient no</td>
<td>Memb cl 1</td>
<td>Memb cl 2</td>
<td>Memb cl 3</td>
<td>Memb cl 4</td>
<td>Memb cl 5</td>
<td>Memb cl 6</td>
</tr>
<tr>
<td>2</td>
<td>1668</td>
<td>0.838108906</td>
<td>0.085469019</td>
<td>0.000203346</td>
<td>0.000123395</td>
<td>0.000123395</td>
<td>0.014881108</td>
</tr>
<tr>
<td>3</td>
<td>1980</td>
<td>0.003581664</td>
<td>0.002136757</td>
<td>0.002504919</td>
<td>0.544640666</td>
<td>1.69E-10</td>
<td>0.004034439</td>
</tr>
<tr>
<td>4</td>
<td>2347</td>
<td>4.16E-06</td>
<td>2.09E-06</td>
<td>2.41E-06</td>
<td>1.69E-10</td>
<td>1.69E-10</td>
<td>0.731640619</td>
</tr>
<tr>
<td>5</td>
<td>3092</td>
<td>0.5</td>
<td>0.085469019</td>
<td>0.5</td>
<td>0.000123395</td>
<td>0.000123395</td>
<td>0.014881108</td>
</tr>
<tr>
<td>6</td>
<td>3675</td>
<td>0.044402543</td>
<td>0.034179609</td>
<td>0.000203346</td>
<td>0.000123395</td>
<td>0.000183125</td>
<td>0.004034439</td>
</tr>
<tr>
<td>7</td>
<td>3810</td>
<td>0.005049351</td>
<td>0.00427351</td>
<td>0.137145964</td>
<td>0.000123395</td>
<td>0.000123395</td>
<td>0.014881108</td>
</tr>
<tr>
<td>8</td>
<td>4096</td>
<td>0.838108906</td>
<td>0.085469019</td>
<td>0.000203346</td>
<td>0.000123395</td>
<td>0.000123395</td>
<td>0.014881108</td>
</tr>
<tr>
<td>9</td>
<td>4097</td>
<td>0.5</td>
<td>0.085469019</td>
<td>0.000203346</td>
<td>0.000123395</td>
<td>0.000123395</td>
<td>0.014881108</td>
</tr>
<tr>
<td>10</td>
<td>4100</td>
<td>0.127328821</td>
<td>0.077449807</td>
<td>0.000203346</td>
<td>0.000123395</td>
<td>0.000123395</td>
<td>0.014881108</td>
</tr>
</tbody>
</table>
Assign a patient to the class with the highest membership function unless...

- The difference between the maximum and the second maximum is less than a threshold
  - If the two are from the same ‘group-family’ then assign the patient to the class with highest membership
  - Otherwise ‘NC’
Current results

Breast Cancer

ER+ CK7/8+

ER- PgR- CK5/6+

HER2+

Not Classified
38 (3.5%)

Luminal

PgR+

PgR-

HER3+ HER4+

HER3- HER4-

Luminal A
Class 1
288 (26.8%)

Luminal N
Class 2
205 (19.1%)

Luminal B
Class 3
186 (17.3%)

Basal

p53+

p53-

HER2/ER+
Class 6A
62 (5.8%)

HER2/ER-
Class 6B
85 (7.9%)

HER3+ HER4+

Luminal N
Class 2
205 (19.1%)

p53 altered
Class 4
113 (10.5%)

p53 normal
Class 5
96 (9.0%)

Luminal A
Class 1
288 (26.8%)

Luminal B
Class 3
186 (17.3%)

new sub-classes

Soria et al., Artificial Intelligence in Medicine, 2013

25

Advanced Data Analysis Centre

28/05/2013
WildTech

- Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife
- EU – FP7 funded project
- Involved in WP7 ‘Database development and WildPro’
  - Provide a mechanism for the integration of data critical to the surveillance of wildlife diseases.
  - Data on samples used to populate the database, including the historical data on the sample and the array screening results.
  - Database data analysed by epidemiologists.

28/05/2013
Major Functions of the System

- Access control of different users
- Upload files of sample spreadsheet
- Upload files of array data
- Export data from the database into files (specific users)
- Query of the database and display the results
- Provide summary of basic statistics of the data
Database structure

shipment
- package_id VARCHAR(30)
- shipper VARCHAR(60)
- ship_contact VARCHAR(45)
- pack_contact VARCHAR(45)
- timestamp TIMESTAMP

indexes

sampledat
- wildtech_sample_id VARCHAR(30)
- package_id VARCHAR(30)
- shipper_sample_id VARCHAR(30)
- shipper_animal_id VARCHAR(30)
- species VARCHAR(30)
- species_if_others VARCHAR(30)
- age VARCHAR(30)
- sex VARCHAR(10)
- weight VARCHAR(30)
- event_id VARCHAR(30)
- other_speciesaffected VARCHAR(30)
- num_of_incidents INT(11)
- lat_lute VARCHAR(30)
- longitude VARCHAR(30)
- reference_system VARCHAR(50)
- country BLOB
- province_state etc BLOB
- region BLOB
- city BLOB
- comments LONGBLOB
- sample_type VARCHAR(30)
- sampled_organ VARCHAR(50)
- ante_postmortem sampling VARCHAR(30)
- animal_status_post_sampling VARCHAR(30)
- type_of_collection VARCHAR(30)
- date_of_death DATETIME
- cause_of_death VARCHAR(50)
- sample_collection_date DATETIME
- sample_condition_at_collection VARCHAR(30)
- sample_storage_date DATETIME

pathother
- index INT(11)
- wildtech_sample_id VARCHAR(30)
- pathogen_other VARCHAR(30)
- infection_status VARCHAR(30)
- type_of_test VARCHAR(30)
- date DATETIME
- test_certification VARCHAR(30)
- lab VARCHAR(30)
- timestamp TIMESTAMP

indexes

arraydata
- index INT(11)
- wildtech_sample_id VARCHAR(30)
- spot_id INT(11)
- partner VARCHAR(30)
- test_date DATETIME
- substance VARCHAR(50)
- confidence DECIMAL(7,6)
- signal DECIMAL(8,6)
- valid INT(11)
- background DECIMAL(7,6)
- mean DECIMAL(7,6)
- test_type VARCHAR(50)
- pathogen VARCHAR(200)
- pos_neg VARCHAR(30)
- timestamp TIMESTAMP

indexes

Advanced Data Analysis Centre

28/05/2013
<table>
<thead>
<tr>
<th>Sample spreadsheet example</th>
<th>Advanced Data Analysis Centre</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WildTech Sample ID</td>
<td>13-000946</td>
</tr>
<tr>
<td>Shipment Sample ID</td>
<td>VLT 1/08</td>
</tr>
<tr>
<td>Shipment Animal ID</td>
<td>Sampled organ</td>
</tr>
<tr>
<td>Species</td>
<td>Moose</td>
</tr>
<tr>
<td>Species if &quot;Other&quot;</td>
<td>Anterior/posterior sampling</td>
</tr>
<tr>
<td>Age</td>
<td>3 years</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
</tr>
<tr>
<td>Date of death</td>
<td>237 kg</td>
</tr>
<tr>
<td>Type of collection</td>
<td>Sample Collection Date</td>
</tr>
<tr>
<td>Sample Condition at Collection</td>
<td>Sample Storage Date</td>
</tr>
<tr>
<td>Sample Collection Date</td>
<td>6650600</td>
</tr>
<tr>
<td>Sample Storage Temperature (°C)</td>
<td>1668000</td>
</tr>
<tr>
<td>Reference System</td>
<td>RT-90</td>
</tr>
<tr>
<td>WildTech Sample Preparation Date</td>
<td>VLT 1/08</td>
</tr>
<tr>
<td>Country</td>
<td>Sweden</td>
</tr>
<tr>
<td>Material left after WildTech Shipment</td>
<td>Amount homogenized/Volume transported</td>
</tr>
<tr>
<td>Region</td>
<td>Stockholm 16th</td>
</tr>
<tr>
<td>Temp of Wildtech Sample Storage at Shipper (°C)</td>
<td>Vld766 Pathological findings</td>
</tr>
<tr>
<td>Test</td>
<td>Frozen</td>
</tr>
<tr>
<td>Tissue</td>
<td>Liver, lung, spleen</td>
</tr>
<tr>
<td>Pathogen (priority)</td>
<td>Mycobacterium bovis</td>
</tr>
<tr>
<td>Infection status</td>
<td>Negative</td>
</tr>
<tr>
<td>Type of test</td>
<td>PCR</td>
</tr>
<tr>
<td>Date</td>
<td>29/08/2011</td>
</tr>
<tr>
<td>Lab</td>
<td>UNOTT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WildTech Sample ID</td>
<td>13-000947</td>
</tr>
<tr>
<td>Shipment Sample ID</td>
<td>VLT 1203/08</td>
</tr>
<tr>
<td>Shipment Animal ID</td>
<td>Sampled organ</td>
</tr>
<tr>
<td>Species</td>
<td>Moose</td>
</tr>
<tr>
<td>Species if &quot;Other&quot;</td>
<td>Anterior/posterior sampling</td>
</tr>
<tr>
<td>Age</td>
<td>1 year</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
</tr>
<tr>
<td>Date of death</td>
<td>Dead</td>
</tr>
<tr>
<td>Type of collection</td>
<td>Sample Collection Date</td>
</tr>
<tr>
<td>Sample Condition at Collection</td>
<td>Sample Storage Date</td>
</tr>
<tr>
<td>Sample Collection Date</td>
<td>4785985</td>
</tr>
<tr>
<td>Sample Storage Temperature (°C)</td>
<td>1473725</td>
</tr>
<tr>
<td>Reference System</td>
<td>RT-90</td>
</tr>
<tr>
<td>WildTech Sample Preparation Date</td>
<td>VLT 1203/08</td>
</tr>
<tr>
<td>Country</td>
<td>Sweden</td>
</tr>
<tr>
<td>Material left after WildTech Shipment</td>
<td>Amount homogenized/Volume transported</td>
</tr>
<tr>
<td>Region</td>
<td>Dalarna</td>
</tr>
<tr>
<td>Temp of Wildtech Sample Storage at Shipper (°C)</td>
<td>Vld76 Pathological findings</td>
</tr>
</tbody>
</table>

28/05/2013
Sample spreadsheet example
Welcome to WildTech

Introduction to WildTech Project

For many reasons, the health of wildlife is of major concern throughout the world. Apart from important influences on the health of many wildlife species, infectious diseases of wildlife have significant impacts on public health and health of livestock. Effective disease surveillance is essential in order to inform control strategies and this depends critically on the development and application of methods of disease diagnosis which are both accurate and rapid. The WildTech project has been established specifically to address these problems and to set up a technology centre that may be exploited in Europe and elsewhere as a basis for high throughput disease diagnosis in wildlife.

The project combines:

- Technological development to enable high throughput nucleic acid- and peptide-based array screening of samples from a wide variety of wild animals
- Surveillance of terrestrial, aerial and marine wild animal species within Europe and from countries which act as portals of disease entry into the EU
- Epidemiological analysis and risk assessment using data generated during the project and from other sources
- Development and proposal of a model framework for disease surveillance within Europe developed in parallel with the existing systems in North America

The project will place the EU at the centre of wildlife disease surveillance and enables the translation of high throughput array-based technologies to human and veterinary medicine.

Click here for the WildTech project website.
Click here for the WildTech probe website.

Last Updated on Monday, 07 November 2011 16:16
Welcome to WildTech

Introduction to WildTech Project

For many reasons, the health of wildlife is of major concern throughout the world. Apart from important influences on the health of many wildlife species, infectious diseases of wildlife have significant impacts on public health and health of livestock. Effective disease surveillance is essential in order to inform control strategies and this depends critically on the development and application of methods of disease diagnosis which are both accurate and rapid. The WildTech project has been established specifically to address these problems and to set up a technology centre that may be exploited in Europe and elsewhere as a basis for high throughput disease diagnosis in wildlife.

The project combines:

- Technological development to enable high throughput nucleic acid- and peptide-based array screening of samples from a wide variety of wild animals
- Surveillance of terrestrial, aerial and marine wild animal species within Europe and from countries which act as portals of disease entry into the EU
- Epidemiological analysis and risk assessment using data generated during the project and from other sources
- Development and proposal of a model framework for disease surveillance within Europe developed in parallel with the existing systems in North America

The project will place the EU at the centre of wildlife disease surveillance and enables the translation of high throughput array-based technologies to human and veterinary medicine.

Click here for the WildTech project website.
Click here for the WildTech probe website.

Last Updated on Monday, 07 November 2011 16:16
Upload Sample

WILDECH DATABASE SYSTEM
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife

MAIN MENU
- Home
- Administrator
- Upload Sample
- Upload Array
- Export Data
- Query Information
- Database Statistics

USER MENU
- Your Details
- Submit an Article
- Submit a Web Link
- Logout

Home > Upload Sample
- Use the "Browse..." button to navigate to the directory where the file is located.
- When the path to the file shows in the text field, click the "Upload File" button.
Upload array

WILDECHT DATABASE SYSTEM
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife

MAIN MENU
- Home
- Administrator
- Upload Sample
- Upload Array
- Export Data
- Query Information
- Database Statistics

USER MENU
- Your Details
- Submit an Article
- Submit a Web Link
- Logout

Home > Upload Array
- Use the “Browse...” button to navigate to the directory where the file is located.
- When the path to the file shows in the text field, tick one of the checkboxes and select the pathogen(s) from the drop-down list (multiple selection allowed).
- When finished, click the “Upload File” button.

Please tick one of the following and choose the pathogen(s):

- Actinobacillus
- AHSSV
- Bacillus
- Bacillus anthracis
- Bluetongue
- Brucella

Positive results for:

- Actinobacillus
- AHSSV
- Bacillus
- Bacillus anthracis
- Bluetongue
- Brucella

Negative results for:

- Actinobacillus
- AHSSV
- Bacillus
- Bacillus anthracis
- Bluetongue
- Brucella

Independent Confirmatory Test (if known):

- Actinobacillus
- AHSSV
- Bacillus
- Bacillus anthracis
- Bluetongue
- Brucella

Positive results for:

- Actinobacillus
- AHSSV
- Bacillus
- Bacillus anthracis
- Bluetongue
- Brucella

Negative results for:

- Actinobacillus
- AHSSV
- Bacillus
- Bacillus anthracis
- Bluetongue
- Brucella

28/05/2013
WILDECH DATABASE SYSTEM
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife

MAIN MENU
- Home
- Administrator
- Upload Sample
- Upload Array
- Export Data
- Query Information
- Database Statistics

USER MENU
- Your Details
- Submit an Article
- Submit a Web Link
- Logout

Home + Export Data
- Please select the name of the table you want to export.
- After the selection, please click "Submit Table" button.

arraydata Submit Table
Export by Barcode

WILDECH DATABASE SYSTEM
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife

MAIN MENU
- Home
- Administrator
- Upload Sample
- Upload Array
- Export Data
- Export By Barcode
- Query Information
- Database Statistics

USER MENU
- Your Details
- Submit an Article
- Submit a Web Link
- Logout

Home > Export By Barcode
- Please insert the barcode you want information for.
- Barcode in the form "xx-xxxxxx" (without inverted commas).
- Example of barcode: 11-000241.
- Then press the Submit button.

Submit
Query information

WILDTech DATABASE SYSTEM
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife

MAIN MENU
- Home
- Administrator
- Upload Sample
- Upload Array
- Export Data
- Query Information
- Database Statistics

USER MENU
- Your Details
- Submit an Article
- Submit a Web Link
- Logout

Home > Query Information
- Please select the name of the table you want to query then click Submit Table.
- Select fields of the table, one for query and the other one for search criterion.
- After two fields are selected, click Submit Fields.
- Search criterion can be typed in the text box, and then click Submit Criterion.

sampledata
Submit Table

wildtech_sample_id
age
Submit Fields

varchar(30)

wildtech_sample_id (age < 6)

- 11-000946
- 11-000947

28/05/2013
WILDTECH DATABASE SYSTEM
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife

MAIN MENU
- Home
- Administrator
- Upload Sample
- Upload Array
- Export Data
- Query Information
- Database Statistics

USER MENU
- Your Details
- Submit an Article
- Submit a Web Link
- Logout

Home → Database Statistics
- Please select a name of the table, then click Submit Table.
- Select a field of the table, and then click Submit Field.

<table>
<thead>
<tr>
<th>sampledst</th>
<th>Submit Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>Submit Field</td>
</tr>
</tbody>
</table>

The average of age is: 2
The minimum of age is: 1 year
The maximum of age is: 3 years
The standard deviation of age is: 1
Alzheimer’s disease

- Panel of molecular markers associated with Alzheimer
- 12 markers scored
  - Blood Plasma – 16 samples
  - Cerebrospinal Fluid (CSF) – 15 samples
- Data sets analysed separately
Methods – Supervised learning

- Model-based classification for prediction of future cases
  - High quality prediction of Healthy / Alzheimer
  - Reduce the number of markers needed
  - Prefer ‘white-box’ prediction model

- Different classification techniques:
  - C4.5 Decision Tree
  - Multi-Layer Perceptron Neural Network
  - Naïve Bayes

- Leave-one-out cross validation used
Summary of results

- Better performance on the Plasma data (more data available)
  - On such small numbers 1 sample can make a difference
- Naïve Bayes never outperformed the other two techniques
  - Can be due to the number of available data
  - This method performs well with big data
- For the CSF, the neural network was the best classifier
  - No classification rules in a readable way
- Interesting results for the C4.5 applied to the Plasma data
  - Knowing the value of just 1 marker (M7) could be enough to classify a patient into the correct group
It is enough to check the values of the M7 variable to establish whether a patient should be classified Healthy or AD.
Save the date!

- ADAC launch event

Wed 2\textsuperscript{nd} October 2013, pm
Thank You!

daniele.soria@nottingham.ac.uk