In an influenza pandemic, identifying individuals whom will subsequently become contagious before having significant exposure with others would enable implementation of public health measures that could limit the spread of disease. The aim of this work was to develop and validate an algorithm for contagiousness using biomarkers identified at early time points post exposure to virus, before subjects would be exhibiting substantial clinical disease symptoms.

To facilitate identification of the prognostic biomarkers we utilised the influenza human viral challenge model (VCM) as it is an effective means of understanding a subject’s pre-exposure biomarker levels and allowing precise monitoring of clinical disease and biomarkers throughout the infection time-course. There are numerous facets to contagiousness. Here we focus on the level of symptoms and viral shedding.

A. Sampling and measures

B. Phenotyping

C. Time-course differential analysis, model selection, algorithm performance

METHODS

We identified numerous biomarkers that had differential time-course trajectories in subjects whom later became contagious from subjects whom did not. By utilising only those biomarkers with the greatest relative influence to minimise the risk of overfitting, we were able to develop several prognostic algorithms that showed good performance within the training set with AUCs from the ROC analysis in the 21 and 45-hour models ranging from 0.75 to 0.87 and 0.72 to 0.75, respectively (Fig. 3).

While the 21-hour model performed slightly better that the 45-hour models in the training data set, all of the 21-hour models performed poorly in independent data set testing. All the 45-hour models, however, performed well in their ability to correctly predict those individuals whom would subsequently become contagious, as indicated by ROC analysis AUC results ranging from 0.82 to 0.89 and optimal accuracy results of 0.85 to 0.88 (Fig. 4).

CONCLUSIONS

- We developed numerous potential prognostic algorithms that were tested on an independent data set.
- The results underscore the importance of independent data-set testing, as not all algorithms retained good performance upon testing.
- 3 different algorithms were shown to retain good predictive performance and therefore passed the independent testing and were validated for contagious prediction and considered ready for field testing. This represents an important step forward in providing additional public health tools for responding to influenza pandemics.
- Furthermore, this work demonstrates the utility of the human challenge model as a powerful tool with which to investigate host response of influenza infection.
- We postulate that the approach described here could also successfully be applied to other disease characteristics.

REFERENCES

The 21-hour models did not perform well on independent data testing, however the 45-hour models demonstrated a high level of predictive performance both for the proteomic (Luminex) and transcriptomic (microarray) analytical platforms as well as the pan-algorithm that combined the two platforms. The independent test data results of the successful algorithms are summarised in the table below, as well as indicating the number of biomarker parameters utilised in each algorithm.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of parameters</th>
<th>ROC curve AUC</th>
<th>ROC curve optimal accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microarray</td>
<td>1</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>Luminex</td>
<td>2</td>
<td>0.82</td>
<td>0.88</td>
</tr>
<tr>
<td>Pan-algorithm</td>
<td>3</td>
<td>0.87</td>
<td>0.85</td>
</tr>
</tbody>
</table>

This work was conducted as part of DARPA’s Prophetic Project and consequently part funded by the US DoD. hVIVO would like to thank DARPA for their collaboration.

ACKNOWLEDGEMENTS

The STRATEGIC APPROACH

Prognostic biomarkers were identified and algorithms developed during the discovery phase via 3 separate phenotype comparisons (Fig. 1). Algorithms were subsequently then tested for contagious prognostic performance in a independent data set.

DISCOVERY PHASE / TRAINING SET (90 subjects)

1. Infected subjects: high vs low viral load
   - Upper tertile PCR vs lower tertile PCR
2. Infected subjects: high vs low viral load
   - Upper tertile VAS vs lower tertile VAS
3. Infected subjects: low/no cold vs uninfected
   - Lower tertile VAS vs uninfected

TESTING PHASE / INDEPENDENT TEST SET (41 subjects)

- Contagious vs Non-contagious
  - Contagious = upper tertile PCR and/or VAS
  - Non-Contagious = remaining subjects

Fig. 1: phenotype comparisons used in both the discovery and testing phases. VAS: visual analogue scale for self reported influenza clinical symptoms.