Using AI to Improve Clinical Specificity of the Karius® Test

Poster Session CIV01—Clinical Studies of Adult Infectious Diseases
Sunday | June 23 | 11:00 AM–1:00 PM

Poster CIV-135
Improving the Clinical Specificity of the Unbiased Karius Test via Literature Mining

Lily Blair1, Desiree Hollemon1, David K. Hong1, Sivan Bercovici1
(1) Karius, Inc., Redwood City, CA

Study Design
Natural language processing and machine learning were applied to process 27 million abstracts from PubMed (December 2018). The association of sepsis-related keywords and microbes was computed using pointwise mutual information (PMI), reflecting how likely a microbe is to cause sepsis. PMI scores were then applied to microbes detected by the Karius Test (clinically adjudicated results) and other microbiological tests in a cohort of 348 patients with suspected sepsis. In the full set of patients, there were 144 positive calls across 140 samples, 52 negative calls from 32 samples, and 234 ambiguous calls.

Results
The PMI scores ranged from -2.3 to 4.9. Among the microbes with the highest scores were Capnocytophaga canimorsus and Fusobacterium nucleatum, and among the lowest were Helicobacter pylori and Alphapapillomavirus. Using a threshold of 0.5 PMI, of the 52 negative calls, 39 calls were classified as not causing sepsis. Of the positives, 141 calls were classified as causing sepsis, leaving only three calls that were confirmed by orthogonal microbiological tests but PMI misclassified as unrelated to sepsis. These include one of Haemophilus haemolyticus (recent literature points to its potential pathogenicity in sepsis) and two of Mycobacterium tuberculosis complex (more commonly associated in the literature with pneumonia).
Improving the Clinical Specificity of the Unbiased Karius® Test via Literature Mining

Lily Blair, Desiree Hollemon, David K. Hong, Sivan Bercovici
Karius, Inc., Redwood City, CA

Background

- Unbiased next generation sequencing (NGS) microbiology tests have led to significant increases in the analytical sensitivity and comprehensiveness, shifting the focus to determining whether the detected microbes are clinically relevant.
- High concentrations of microbial cell-free DNA (mcfDNA) have been associated with confirmed infectious disease diagnosis. However, high concentration is not sufficient to identify the cause of infection in all cases.
- We developed a method for establishing associations between microbes with specific diseases by mining the literature, resulting in the improved clinical specificity of the Karius Test.

Methods

- Natural language processing was applied to process more than 27 million abstracts from PubMed (December 2018).
- Association of sepsis-related keywords and microbes was computed using pointwise mutual information (PMI), reflecting how likely a microbe is to cause sepsis, given its presence.
- PMI scores were calculated for 1320 microbes in the clinically reportable range of the Karius Test.
- PMI scores were applied to microbes detected in a cohort of 348 patients, who presented to the Stanford Hospital emergency department with sepsis alert 1.
- Microbes identified by the Karius Test were compared to microbiological testing or adjudicated as causing sepsis and also are common commensals, such as Haemophilus influenzae and Streptococcus mitis.
- All incorrectly classified negatives were associated with patients that were discharged due to suspected viral infection not covered by the Karius test.
- Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease.
- Microbiome sequencing results were adjudicated as the cause of sepsis by a panel of three independent infectious disease doctors 2.

Scores from Literature Mining

- A median of 34 abstracts mentioned each microbe, with E. coli described in 198,513.
- 788 microbes that never co-occurred with sepsis keywords resulted in undefined PMI scores and were assigned a PMI score of 0.
- Capnocytophaga canimorsus and Fusobacterium necrophorum are among the microbes with the highest PMI scores associating them to sepsis.
- Alphapapillomavirus and Helicobacter pylori are among the microbes with the lowest PMI scores.
- We note that some disease-microbe associations are poorly captured in the literature. Furthermore, PMI scores can be biased toward lower scores for microbes that are model organisms, which are very common in the literature.

Classification based on literature mining scores improves specificity

- PMI scores, with a threshold of 0.5, were used to classify microbes as likely causing or not causing sepsis.
- Positives are microbes that were confirmed by microbiology testing or adjudicated as causing sepsis.
- Negatives are microbes detected by the Karius Test alone and adjudicated as an unlikely cause of sepsis.
- Of 144 positives, 141 were classified as causing sepsis.
- Of 52 negatives, 39 were classified as not causing sepsis.
- The negatives that were incorrectly classified as causing sepsis were microbes that commonly cause sepsis and also are common commensals, such as Haemophilus influenzae and Streptococcus mitis.
- Literature mining correctly identifies microbes that are relevant to sepsis and is easily applied to additional clinical indications.

Conclusions

- Literature mining correctly identifies microbes that are relevant to sepsis and is easily applied to additional clinical indications.
- Literature mining provides a powerful tool aiding in the interpretation of unbiased microbial NGS tests.

1. The SIS-SEQ study (NCT02730468)