Peripheral blood glycoproteomic biomarkers as a powerful new tool for the detection of nasopharyngeal carcinoma (NPC)

Aye TT[1]; Xu G[2]; Mitchell A[2]; Serie D[2]; Malek A[1]; Low LH[1]; Tan LP[3]; Low CT[3]; Tan KE[3]; Wong CY[4]; Sawali H[6] ; Yap YY[6] ; Khoo AS[3,7] ; Lindpaintner K[2]

[1]InterVenn Biosciences, Malaysia; [2]InterVenn Biosciences, South San Francisco, CA; [3]Institute of Medical Research, Malaysia; [4]Department of Otorhinolaryngology Sarawak General Hospital, Malaysia ; [5]Department of Otorhinolaryngology, Queen Elizabeth Hospital, Malaysia; [6]Department of Otorhinolaryngology, Kuala Lumpur Hospital, Malaysia; [7]International Medical University, Malaysia

Introduction

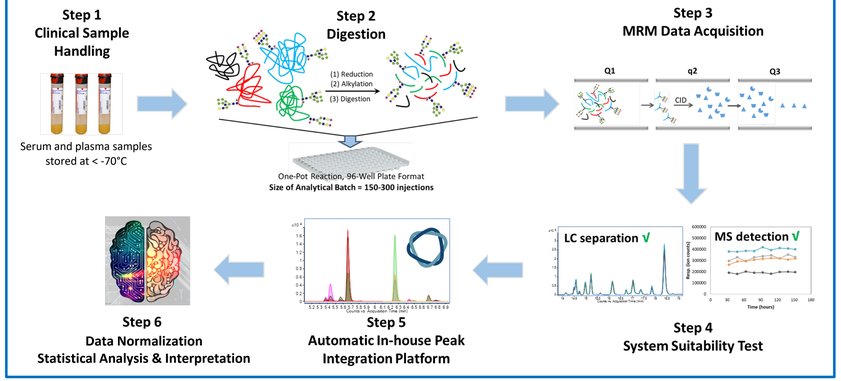
Nasopharyngeal cancer (NPC) is a rare type of head and neck cancer that occurs at the nasopharynx, with the highest incidence in east and Southeast Asia. The incidence rate for NPC is 5-30 cases per 10000 person-years with average overall survival rate ranges from 50-90% depending on the disease stages. As NPC remains largely asymptomatic at early stages, most NPC patients are diagnosed at the advanced stage. Hence, accurate and effective methods are in urgent need for the early diagnosis, evaluation or prognosis for NPC.

While Epstein-Barr virus (EBV) DNA and EBV serology test have been advocated for the early diagnosis and screening for NPC, the sensitivity of these two tests vary because only a small proportion of individuals infected with EBV virus would ultimately develop NPC.

Aberrant glycosylation of proteins play a critical roles in tumor development and we have recently demonstrated that interrogating the glycoproteome, using a proprietary platform that couples artificial intelligence (AI) to targeted liquid chromatography-mass spectrometry (LC-MS) yields highly informative biomarkers in a range of neoplastic diseases.

Methods

This proof-of-concept study consisted of 106 plasma samples (1:1 ratio of histologically confirmed cases of NPC and cancer-free control) which were matched of age (average 49 years). All subjects were male from a multi-ethnic patient population in Malaysia. A panel of >500 glycopeptides was selected for high throughput analysis and further subjected to multivariable statistical analysis.



Sample Information

Blood specimens were collected by the Cancer Research Center, Institute of Medical Research, Ministry of Health, Malaysia. Each sample was staged based on the histopathological report. To control for the effect of genetic variants, the plasma samples from four ethnic groups were included (Table 1a and 1b). The study were conducted in accordance with the international conference on Harmonization guidelines for Good Clinical Practice (ICH E6) and the code of Federal regulation on the protection of Human Subject (45 CFR Part 46)

Table 1a Clinical profiles and demographics of healthy controls and NPC patients

Classification	Description	Cases	Control
Age	Range	(40,58)	(39,61)
	Mean ± SD	49.8 ± 4.8	49.9 ± 5.8
Gender	Male	Male	Male
Ethinicity	Chinese	19	19
	Kadazan/Kadazan Dusun	12	12
	Iban	6	6
	Bidayuh	16	16

Table 1b Clinical stages where molecular sub-typed of stages of NPC cancer patients were classified

no. cases	American Joint Committee on Cancer (AJCC) Staging, 7th edition
0	I
9	II
20	III
6	IVA
15	IVB
3	IVC

Results and Discussion

Among a panel of > 500 glycopeptides and peptides targeted, 62 glycopeptides were differentially expressed between NPC cases and healthy controls. Figure 1 showed a heatmap of the relative abundance of the 36 targeted markers which were significantly regulated in NPC cases (FDR<0.001).

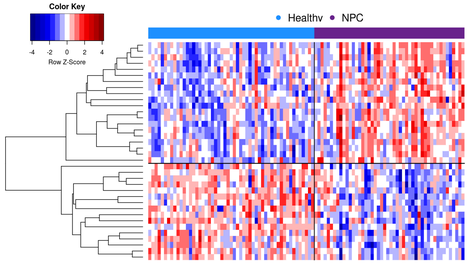


Figure 1. 36 biomarkers show statically significantly between healthy control and cancer cases

Multivariate ML Modelling

A multivariable leave-one-out repeated cross-validated **LASSO-regularized logistic regression model** predicting probability of NPC cases from healthy control. It shows that high specificity and sensitivity were achieved with an AUC of 0.955 for test samples whereas 0.996 for the training set, respectively.

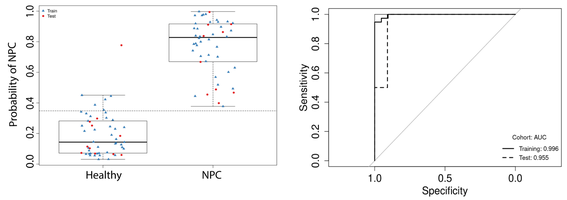


Figure 2. Box plot and ROC curves to display the power of differentially expressed glycopeptides and aglycopeptides in NPC cases and healthy control

Staging of NPC using 5 selected biomarkers

Five glycopeptide were selected to differentiate the stages of disease progressiveness in NPC.

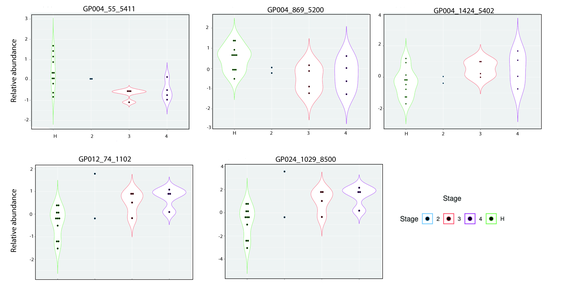


Figure 2. Violin plot display the distribution of specific marker across the different stages of NPC whereas H represents Healthy control, 2,3, and 4 represent the stages of the NPC cases.

- Glycopeptides levels **clearly differentiate NPC cases from healthy controls.**
- A subset of 62 biomarkers were shown **specific to NPC cases**, potentially providing insights into specific biomarkers for early diagnosis of NPC.
- Five markers were shown strong discriminating power in **differentiating the stages of cancer**.
- **Low heterogeneity** observed across a variety of ethnicities
- **Follow-up studies** with larger sample size are currently in progress

Conclusions

Glycoproteomic profiles are statistically significantly different from NPC patients and healthy controls. InterVenn's proprietary perspectIV™ platform, based on a high-throughput LC-MS/AI-ML approach, interrogates the blood glycoproteome at an unprecedented depth and scale, revealing a whole new domain of biology for the identification of novel diagnosis biomarkers in NPC.