Review Article

Metabolic Biomarkers of Squamous Cell Carcinoma of the Aerodigestive Tract: A Systematic Review and Quality Assessment

Yan Mei Goh, Stefan S. Antonowicz, Piers Boshier, and George B. Hanna

Department of Surgery & Cancer, Imperial College London, London W2 1NY, UK

Correspondence should be addressed to George B. Hanna; g.hanna@imperial.ac.uk

Received 5 September 2019; Revised 22 December 2019; Accepted 21 January 2020; Published 21 February 2020

Guest Editor: Kanhaiya Singh

Copyright © 2020 Yan Mei Goh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction. Aerodigestive squamous cell carcinomas (ASCC) constitute a major source of global cancer deaths. Patients typically present with advanced, incurable disease, so new means of detecting early disease are a research priority. Metabolite quantitation is amenable to point-of-care analysis and can be performed in ASCC surrogates such as breath and saliva. The purpose of this systematic review is to summarise progress of ASCC metabolomic studies, with an emphasis on the critical appraisal of methodological quality and reporting.

Method. A systematic online literature search was performed to identify studies reporting metabolic biomarkers of ASCC. This review was conducted in accordance with the recommendations of the Cochrane Library and MOOSE guidelines.

Results. Thirty studies comprising 2117 patients were included in the review. All publications represented phase-I biomarker discovery studies, and none validated their findings in an independent cohort. There was heterogeneity in study design and methodological and reporting quality. Sensitivities and specificities were higher in oesophageal and head and neck squamous cell carcinomas compared to those in lung squamous cell carcinoma. The metabolic phenotypes of these cancers, as was the kinetics of metabolite groups when comparing blood, tissue, and breath/saliva concentrations. Deregulation of amino acid metabolism was the most frequently reported theme.

Conclusion. Metabolite analysis has shown promising diagnostic performance, especially for oesophageal and head and neck ASCC subtypes, which are phenotypically similar. However, shortcomings in study design have led to inconsistencies between studies. To support future studies and ultimately clinical adoption, these limitations are discussed.

1. Introduction

Squamous cell carcinomas of the aerodigestive tract (ASCC) constitute a major health burden globally, with an estimated 4.3 million new cases and 2.6 million deaths annually [1]. Poor survival that is associated with ASCC reflects their often delayed presentation to medical professionals, such that many patients are not suitable for curative therapy [2–5]. Whilst the ability to diagnose ASCC at an early stage is associated with improved long-term survival, current strategies have inadequate diagnostic performance and are not recommended in national guidelines. There remains an unmet clinical need to develop reliable noninvasive and cost-effective methods for the early detection of ASCC.

ASCC arise from nonkeratinising stratified squamous epithelium lining the upper digestive tract (lips to lower oesophagus) and respiratory tract. This convenient location renders ASCC suitable for noninvasive testing using breath and saliva. The use of proteomics and genomics has historically been at the forefront of diagnostic studies. However, these techniques provide monothematic information and are less suited to point-of-care technologies needed for large-scale application [6]. Metabolites may be more appealing as they are amenable to noninvasive sampling and translatable to point-of-care analytical tools [7]. For example, in upper gastrointestinal adenocarcinoma (the other major ASCC subtype), exhaled metabolites have demonstrated promise for detecting treatable disease stages [8–10]. However, progress in this field has been hampered by inadequate standardisation, inconsistent quality assurance, and evolving analytical technology [11–14].

The purpose of this systematic review is to summarise progress of ASCC metabolomic studies. The specific objectives are (i) to assess methodological quality, (ii) to summarise the
discriminatory performance of the proposed metabolic biomarkers, and (iii) to describe emerging metabolic themes for these cancers.

2. Materials and Methods

2.1. Literature Search. This review set out to identify all studies that measured differences in metabolites between patients with ASCC and relevant controls. A systematic literature search was conducted in accordance with the recommendations of the Cochrane Library and MOOSE guidelines [15]. The following databases were searched: Medline (1946–present) via OvidSP, Ovid Embase (1947–18th January 2019), and Cochrane Library. Three strings using the following search terms were used: biomarkers; metabolomics; metabolic profiling; volatile organic compounds; magnetic resonance spectroscopy; mass spectrometry; and squamous cell carcinoma. All variations in spelling including a truncated search term were searched. The search was conducted in accordance with the recommendations of the Cochrane Library and MOOSE guidelines [15].

2.2. Definitions. Metabolomics is defined as “the global and unbiased definition of the complement of small molecules in biofluids, tissues, organs, or organisms” [16]. Biomarkers were defined as naturally occurring molecules, which were significantly different in a disease state. ASCC included tumours affecting squamous mucosa of the oral cavity, oropharynx, lung, and oesophagus.

2.3. Outcome Measures. The following data items were extracted from included publications: year of publication, country of origin, study design, recruitment time, total number of participants, tumour of origin, biomarker phase, tumour stage, analytical platform used, sample type, number of compounds identified, compounds noted to be increased/decreased in cancer, statistical analysis performed, prediction model used, sensitivity and specificity, and area under the receiver operating characteristic (ROC) curve derived from diagnostic models.

2.4. Statistical Analysis. Statistical analysis was performed using R (version 3.2.1, The R Project for Statistical Computing, http://www.r-project.org). Using the sensitivity, specificity, and area under the ROC curves derived from individual published models, bivariate meta-analyses were performed to create pooled point estimates of the hierarchical summary ROC curve of VOC analysis in accordance with previously validated methods [17].

2.5. Metabolite Analysis. All metabolites identified were classed in accordance to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and statistical analysis was performed using the pathway analysis module in MetaboAnalyst version 4.0. Metabolites determined to be significantly increased or decreased in each study were selected. Data pre-processing included name check against the Human Metabolome Database (HMDB), data checks, and missing values. Parameters used to analyse this data were the hypergeometric test for overrepresentation analysis and the relative-betweenness centrality test for pathway topology analysis based on the KEGG pathway library [18–20]. Normalisation was performed using the weighted means of identified metabolite that were increased/decreased in squamous cell carcinoma (SCC) in each sample type. The mean proportion of each compound identified was assessed as the proportion of the total number of compounds identified per metabolite class per study, divided by the total number of compounds identified in total in each SCC site subtype, multiplied by the total number of studies; this compound was identified in Figure 1.

2.6. Quality Assessment. Study quality was assessed with three tools: first, Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist [21] to assess methodological bias. Second, the Standards for Reporting of Diagnostic Accuracy Studies (STARD) checklist [22, 23] was used to assess general reporting quality of a clinical diagnostic tool. Third, the Chemical Analysis Working Group- (CAWG-) Metabolomics Standard Initiative (MSI) criteria was used as this focused on the reporting quality of metadata of metabolomic studies [14]. The CAWG-MSI Metabolite Identification Levels were used to summarise studies’ identification rigour: level 1 (most confident, at least two orthogonal analytical data types, e.g., retention time, isotope labelling), level 2 (one data type, spectral similarity to commercial library), level 3 (one data type related to a spectral or chemical property).

3. Results

A systematic literature search identified a total of 30 studies comprising of a total of 2117 subjects of which 1144 had a diagnosis of ASCC (Figure 2). Details of included studies were provided in Table 1. All studies were Phase I biomarker discovery studies. Of the 30 included studies, 18 were from Asia and the Far East [8, 9, 11–26], seven from Europe [27–33], three from North America [34–36], and two from the Middle East [37, 38]. ASCC tumour sites identified were the head and neck (n = 17), oesophageal (n = 8), and lung (n = 5). The majority of studies compared patients with cancer to normal controls or benign conditions [16, 25, 26, 28, 29, 34, 38–43].

Liquid chromatography mass spectrometry (LC-MS) was the most commonly used analytical platform (n = 14)
followed by gas chromatography mass spectrometry (GC-MS, \( n = 12 \)). Sample types used in these studies were tissue \( (n = 10) \), saliva \( (n = 13) \), plasma \( (n = 13) \), urine \( (n = 5) \), and breath \( (n = 6) \). Several studies used more than one analytical platform \([34, 35, 38, 40]\) and/or sample types for analysis \([34, 35]\) (Table 1). Eight studies used targeted methods, and 22 studies untargeted methods. All studies that used untargeted methods covered a large range of commonly identified metabolites, ranging from small fatty acids to larger glycolipid and carbohydrate metabolites. Only five studies identified volatile compounds \([26, 30, 36–38]\).

### 3.1. Quality

Assessment of bias and applicability of outcomes were analysed with QUADAS-2 (Table 1). The QUADAS-2 was divided into risk of bias of the following: patient selection, diagnostic test, reference standard, and patient flow and timing. Additionally, this test investigated applicability of patient selection, diagnostic test, and
<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Sample type</th>
<th>SCC stage</th>
<th>Targeted/untargeted method</th>
<th>Analytical platform</th>
<th>Statistical analysis/prediction model</th>
<th>STARD score</th>
<th>QUADAS Risk of bias</th>
<th>Applicability</th>
<th>CAWG-MSI metabolite ID level</th>
<th>CAWG-MSI score Sn (%)</th>
<th>CAWG-MSI score Sp (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu 2013</td>
<td>China</td>
<td>Plasma</td>
<td>Late: 17</td>
<td>Untargeted</td>
<td>UPLC-ESI-TOF-MS</td>
<td>PCA, hierarchical cluster analysis</td>
<td>37</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Wang L 2013</td>
<td>China</td>
<td>Tissue</td>
<td>Early: 28</td>
<td>Untargeted</td>
<td>IH-NMR</td>
<td>OPLS-DA</td>
<td>35</td>
<td>Low</td>
<td>High</td>
<td>2</td>
<td>7</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Jin 2014</td>
<td>China</td>
<td>Plasma</td>
<td>Early: 49</td>
<td>Untargeted</td>
<td>GC-MS</td>
<td>Model of 3 compounds based on OPLS-DA model</td>
<td>33</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>16</td>
<td>90</td>
<td>0.964</td>
</tr>
<tr>
<td>Ma 2014</td>
<td>China</td>
<td>Plasma</td>
<td>Early: 51</td>
<td>Targeted</td>
<td>HPLC</td>
<td>Student t-test, PLS-DA</td>
<td>32</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>10</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Wang J 2016</td>
<td>China</td>
<td>Plasma</td>
<td>Early: 28</td>
<td>Untargeted</td>
<td>UHPLC-QTOF/MS</td>
<td>Model of 16 compounds based on random forest model</td>
<td>37</td>
<td>Low</td>
<td>Low</td>
<td>1</td>
<td>19</td>
<td>85</td>
<td>90.5</td>
</tr>
<tr>
<td>Xu 2016</td>
<td>China</td>
<td>Urine</td>
<td>Late: 40</td>
<td>Untargeted</td>
<td>LC-MS/MS</td>
<td>Model of 7 compounds based on binary logistic regression and ROC curve</td>
<td>25</td>
<td>Unclear</td>
<td>Low</td>
<td>2</td>
<td>19</td>
<td>90.2</td>
<td>96.0</td>
</tr>
<tr>
<td>Cheng 2017</td>
<td>China</td>
<td>Plasma</td>
<td>Patient: 40</td>
<td>Targeted</td>
<td>LC-MS/MS</td>
<td>Model of 4 compounds based on fivefold cross-validation test</td>
<td>32</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>15</td>
<td>77.5</td>
<td>85.33</td>
</tr>
<tr>
<td>Zhang 2017</td>
<td>China</td>
<td>Plasma</td>
<td>Early: 17</td>
<td>Untargeted</td>
<td>1H-NMR, UHPLC</td>
<td>Model of 9 compounds based on binary logistic regression and ROC curves</td>
<td>33</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>14</td>
<td>97.4</td>
<td>95</td>
</tr>
</tbody>
</table>

### Studies of lung squamous cell carcinoma

<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Sample type</th>
<th>SCC stage</th>
<th>Targeted/untargeted method</th>
<th>Analytical platform</th>
<th>Statistical analysis/prediction model</th>
<th>STARD score</th>
<th>QUADAS Risk of bias</th>
<th>Applicability</th>
<th>CAWG-MSI metabolite ID level</th>
<th>CAWG-MSI score Sn (%)</th>
<th>CAWG-MSI score Sp (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song 2010</td>
<td>China</td>
<td>Breath</td>
<td>Early: 20</td>
<td>Untargeted</td>
<td>SPME GC-MS</td>
<td>Wilcoxon rank sum test, ROC</td>
<td>29</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>11</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>De Castro 2014</td>
<td>Spain</td>
<td>Plasma</td>
<td>Patient: 30</td>
<td>Targeted</td>
<td>GC-MS</td>
<td>Model of 1 compound</td>
<td>30</td>
<td>Unclear</td>
<td>Low</td>
<td>1</td>
<td>13</td>
<td>77</td>
<td>66</td>
</tr>
<tr>
<td>Author</td>
<td>Country</td>
<td>Sample type</td>
<td>SCC stage</td>
<td>Targeted/untargeted method</td>
<td>Analytical platform</td>
<td>Statistical analysis/prediction model</td>
<td>STARD score</td>
<td>QUADAS Risk of bias</td>
<td>CAWG-MSI metabolite ID level</td>
<td>CAWG-MSI score</td>
<td>Sn (%)</td>
<td>Sp (%)</td>
<td>AUC</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
<td>--------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>Handa 2014</td>
<td>Germany</td>
<td>Breath</td>
<td>Early: 19 late: 31 Normal: 39</td>
<td>Untargeted</td>
<td>IMS</td>
<td>Model of 11 compounds based on decision tree algorithm</td>
<td>27</td>
<td>Unclear</td>
<td>Low</td>
<td>3</td>
<td>6</td>
<td>97.4</td>
<td>60</td>
</tr>
<tr>
<td>Rocha 2015</td>
<td>Portugal</td>
<td>Tissue</td>
<td>Patient: 19 Control: 37</td>
<td>Untargeted</td>
<td>I1H-NMR</td>
<td>PLS-DA, Wilcoxon rank sum test</td>
<td>25</td>
<td>Unclear</td>
<td>Low</td>
<td>2</td>
<td>7</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sanchez-Rodriguez 2015</td>
<td>Spain</td>
<td>Plasma</td>
<td>Late: 18 Control: 50</td>
<td>Targeted</td>
<td>GC-MS</td>
<td>Model of 1 compound based on ROC curves</td>
<td>31</td>
<td>Low</td>
<td>Low</td>
<td>1</td>
<td>17</td>
<td>69</td>
<td>68</td>
</tr>
<tr>
<td>Mizukawa 1998</td>
<td>Japan</td>
<td>Saliva</td>
<td>Patient: 18 Control: 18</td>
<td>Targeted</td>
<td>HPLC</td>
<td>Nil-peak detection only</td>
<td>21</td>
<td>Low</td>
<td>High</td>
<td>1</td>
<td>7</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Somashekar 2011</td>
<td>USA</td>
<td>Tissue</td>
<td>Patient: 22 Control: 22</td>
<td>Untargeted</td>
<td>HR-magic angle spinning proton NMR spectroscopy</td>
<td>23</td>
<td>Low</td>
<td>Low</td>
<td>1</td>
<td>8</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Wei 2011</td>
<td>China</td>
<td>Saliva</td>
<td>Early: 21 Late: 16 Control: 66</td>
<td>Untargeted</td>
<td>UPLC-QTOF-MS</td>
<td>Model of 5 compounds based on ROC curves</td>
<td>31</td>
<td>Low</td>
<td>Low</td>
<td>3</td>
<td>13</td>
<td>86.5</td>
<td>82.4</td>
</tr>
<tr>
<td>Yonezawa 2013</td>
<td>Japan</td>
<td>Tissue, plasma</td>
<td>Early: 7 Late: 10 Control: 22</td>
<td>Untargeted</td>
<td>GC-MS</td>
<td>Student’s t-test, Bartlett’s test, Wilcoxon rank sum test</td>
<td>27</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>17</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gruber 2014</td>
<td>Israel</td>
<td>Breath</td>
<td>Early: 9 Late: 11 Control: 40</td>
<td>Untargeted</td>
<td>GC-MS, sensors</td>
<td>Model of 3 compounds based on discriminant factor analysis</td>
<td>30</td>
<td>Low</td>
<td>Low</td>
<td>3</td>
<td>10</td>
<td>77</td>
<td>90</td>
</tr>
<tr>
<td>Wang Q (Clinica Chimica Acta) 2014</td>
<td>China</td>
<td>Saliva</td>
<td>Early: 13 Late: 17 Control: 0</td>
<td>Targeted</td>
<td>UPLC-MS</td>
<td>Model of 4 compounds based on ROC curves</td>
<td>30</td>
<td>Unclear</td>
<td>Low</td>
<td>1</td>
<td>24</td>
<td>92.3</td>
<td>91.7</td>
</tr>
<tr>
<td>Wang Q (Scientific Reports) 2014</td>
<td>China</td>
<td>Saliva</td>
<td>Early: 13 Late: 17 Control: 30</td>
<td>Untargeted</td>
<td>RPLC-MS, HILIC-MS</td>
<td>Model of 5 compounds based on ROC curve</td>
<td>24</td>
<td>Unclear</td>
<td>Low</td>
<td>1</td>
<td>16</td>
<td>100</td>
<td>96.7</td>
</tr>
<tr>
<td>Author</td>
<td>Country</td>
<td>Sample type</td>
<td>SCC stage</td>
<td>Targeted/untargeted method</td>
<td>Analytical platform</td>
<td>Statistical analysis/prediction model</td>
<td>STARD score</td>
<td>QUADAS Risk of bias</td>
<td>Applicability</td>
<td>CAWG-MSI metabolite ID level</td>
<td>CAWG-MSI score</td>
<td>Sn (%)</td>
<td>Sp (%)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>----------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Wang Q (Talanta) 2014</td>
<td>China</td>
<td>Saliva</td>
<td>Early: 13</td>
<td>Targeted</td>
<td>UPLC-ESI-MS</td>
<td>Model of 2 compounds based on logistic regression model</td>
<td>25</td>
<td>Low</td>
<td>Unclear</td>
<td>1</td>
<td>25</td>
<td>92.3</td>
<td>91.7</td>
</tr>
<tr>
<td>Gupta 2015</td>
<td>India</td>
<td>Plasma</td>
<td>Early: 28</td>
<td>Untargeted</td>
<td>H-NMR</td>
<td>Model of 2 compounds based on OPLS-DA</td>
<td>33</td>
<td>Unclear</td>
<td>Low</td>
<td>2</td>
<td>10</td>
<td>90</td>
<td>94</td>
</tr>
<tr>
<td>Szabo 2015</td>
<td>Hungary</td>
<td>Breath</td>
<td>Cancer: 14</td>
<td>Targeted</td>
<td>OralChroma and GC-MS</td>
<td>Nil–peak detection only Kruskal–Wallis, Fisher exact test, Cox proportional hazards model</td>
<td>22</td>
<td>Unclear</td>
<td>Low</td>
<td>1</td>
<td>8</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mukherjee 2016</td>
<td>USA</td>
<td>Tissue, saliva</td>
<td>Early: 2</td>
<td>Untargeted</td>
<td>LC-MS, LC-MS/MS, GC-MS</td>
<td>Kruskal–Wallis with adjustment for multiple testing</td>
<td>36</td>
<td>Low</td>
<td>Low</td>
<td>3</td>
<td>15</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Hartwig 2017</td>
<td>Germany</td>
<td>Breath</td>
<td>Early: 5</td>
<td>Untargeted</td>
<td>GC-MS</td>
<td>Jackknife/leave-one-out cross-validation</td>
<td>34</td>
<td>Unclear</td>
<td>Low</td>
<td>3</td>
<td>6</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Kamarajan 2017</td>
<td>USA</td>
<td>Tissue, saliva, plasma</td>
<td>Early: 17</td>
<td>Untargeted</td>
<td>UPLC-MS/MS, GC-MS</td>
<td>Anova, t-test, random forest classification, PCA</td>
<td>31</td>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>20</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ohshima 2017</td>
<td>Japan</td>
<td>Saliva</td>
<td>Early: 14</td>
<td>Untargeted</td>
<td>CE-TOF-MS</td>
<td>Hierarchical cluster analysis, Wilcoxon rank sum test</td>
<td>37</td>
<td>Low</td>
<td>Low</td>
<td>3</td>
<td>9</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

reference standard to the systematic review. There was an overall low risk of bias of these diagnostic tests and high applicability of these studies to the review question. In this QUADAS-2 analysis, the nature of patient flow and timing of sample analysis was least reported in studies in this review \((n = 11)\) \[16, 28–30, 33, 36, 39, 40, 44\].

General reporting quality of a clinical diagnostic tool was assessed by the STARD checklist (Table 1). The STARD score for reported studies ranged from 21 to 37 with a mean of 29.4 \(\pm 4.76\) S.D where the maximum score is 41. More than 75% of studies reported inclusion and exclusion criteria, described the reference test and standards, and reported potential bias and analysis of diagnostic accuracy well. However, more than two thirds of studies failed to clearly demonstrate patient recruitment protocol, specifically, how patients were identified and recruited, the nature of recruitment, e.g., consecutive or random series and [22, 37, 40, 41, 45–47] sample size estimation [35, 42, 48], participant flow [24, 31, 35, 39, 49, 50], and adverse effects as a consequence of the diagnostic tool.

Reporting of clinical demographics was not consistent in each study. Of the 30 studies, only 10 fully reported all clinical demographics [28, 30, 34, 35, 38, 41, 43, 47, 49, 51], 13 reported at least patient age, gender, and clinical stage [16, 25, 26, 31, 36, 39, 40, 42, 45, 46, 48, 52, 53]. Seven studies did not report differences in metabolite profile at different tumour stages [24, 27, 29, 32, 36, 37, 46]. In these seven studies, four compared differences in metabolic profile between cancer and non-cancer cohorts [24, 27, 32, 37]. Definitions of normal control differed most in tissue sample analysis, where Zhang et al. specified normal adjacent control tissue samples a minimum of 5 cm from the tumour site [54] in contrast to the other five tissue studies that used adjacent normal controls [26, 29, 45, 46, 53] without demonstrating adequacy of tissue clearance. Of all 30 studies in this review, only Shoffel-Havukak et al. used patients with benign histology as controls [37]. No tissue study used normal samples from patients with no endoluminal pathology, which is pertinent as metabolic field effects exist in endolumens [55]. Various exclusion criteria were given to control donors’ characteristics, including use of nonsteroidal anti-inflammatory drugs within the past week, antibiotic treatment and consumption of specific food, history of mucosal disorder, chronic and/or systemic disease such as diabetes, autoimmune disorders, heart disease, infection, and liver disease. Twenty-six studies involving biofluids or breath used healthy volunteer controls, one additionally used patients with benign diseases [24, 25, 27, 28, 30–43, 46–49, 51, 52, 56, 57]. The definition for healthy volunteers was based on history (six studies) or endoluminal study (18 studies).

Reporting of metadata in metabolomics datasets was assessed using CAWG-MSI [14] (Supplementary Table 2). A summary of the minimum reported metadata is summarised in Table 1. Twenty of the 30 studies included in this systematic review used relative quantification of compounds [16, 24, 25, 28, 29, 31, 33–38, 40–43, 45, 48, 49, 52], whilst 10 included studies provided absolute quantification of compounds [27, 30, 32, 39, 46, 47, 51–53, 56]. Despite the availability of reporting guidelines for metabolomics analysis, only three studies reported greater than 50% of the CAWG-MSI criteria [34, 39, 52]. Overall, studies reported sample preparation, experimental analysis, and instrumental performance well. However, the majority (80%) did not provide method validation data [16, 24, 25, 27, 28, 30–38, 40–42, 45–49, 51, 52, 54, 56]. Thirteen studies that analysed relative quantification of metabolites identified used either internal standards or normalised the results to allow for instrument variation [16, 24, 25, 29, 33–36, 40–42, 45, 49]. Six of the 10 studies that used absolute quantification did not report accuracy or precision validation data for their method on the instrument [30, 44, 46, 47, 51, 56] whilst two of 10 studies reported the limits of quantification and detection of their method [39, 52]. Out of 30 studies, only 12 declared evidence of data preprocessing [25, 29, 33, 34, 36, 37, 40, 41, 43, 45, 48, 49]. Levels one, two, and three metabolite identification were reported in nine [25, 27, 30, 32, 36, 39, 40, 46, 52], 15 [16, 24, 26, 29, 33, 34, 37, 42, 43, 45, 47, 49, 51, 56, 58], and six [28, 31, 35, 38, 41, 48] studies, respectively. Only two of the 30 studies reported all of the statistical aspects suggested by the CAWG-MSI guidelines [59].

3.2. Discriminatory Features. The highest sensitivity of oesophageal squamous cell cancer (OSCC) diagnosis was reported by Zhang et al. at 97.4% with a specificity of 95% and AUC of 0.988 [44]. Jin et al. reported the highest specificity at 96.67% with a sensitivity of 90% and AUC of 0.964 [42]. The highest sensitivities and specificities of lung squamous cell cancer (LSCC) were poorer with Handa et al. reporting the highest sensitivity of 97.4% [28] and Sanchez-Rodriguez et al. reporting the highest specificity of 68% and AUC of 0.7 [30]. The highest sensitivity, specificity, and AUC were reported for head and neck squamous cell cancer (HNSCC): 100%, 96.7%, and 0.997, respectively (Table 1). However, no groups subsequently validated their initial findings in independent cohorts. Of the 6 studies which reported AUC > 0.90, a high risk of bias was not present and CAWG-MSI metabolite identification was level 1 or 2.

3.3. Metabolic Themes. A total of 181 metabolites identified were associated with an increase or decrease in concentration in patients with ASCC compared to their normal controls (Supplementary Table 3). These compounds were identified in a range of sample types including tissue, plasma, urine, saliva, and breath. The majority were amino acids, carboxylic acids, or fatty acids, and these were more commonly identified in tissue, saliva, and plasma samples. The least common metabolites identified were vitamins, nitrogen, and sulphur containing compounds (Supplementary Figure 1). Sixty-eight compounds that changed in ASCC were reported in more than one study. These metabolites were selected based on metabolites that were identified to be increased or decreased in cancer in different studies. Of these, 27 compounds were noted to be involved in amino acid and lipid metabolism (Supplementary Table 4). All biomarkers showed a consistent increase or decrease in the sample types across different studies (see Supplementary Tables 3 and 4).
A particularly deregulated pathway was branched chain amino acid metabolism (BCAAs, see Figures 3 and 4 and Supplementary Table 3). There were 36 significant differences in BCAA concentrations, or their downstream metabolites, across 12 studies [25, 28, 34–36, 39–41, 43, 45, 46, 49]. The QUADAS-2 risk of bias was low for nine of these studies [34–36, 40, 41, 45–47, 49]. Five of these studies were of good quality and five of fair quality as assessed by the STARD checklist. One study reported the minimum metadata required from the CAWG-MSI checklist [39]. Of these 12 studies, two reported level 1 metabolite identification [35, 39], seven reported level 2 [25, 28, 36, 41, 43, 45, 60], and three reported level 3 metabolite identification [34, 40, 46].

3.4. Influence of Anatomical Location on Metabolic Themes. LSCC (n = 9) had the lowest number of metabolite classes compared to OSCC (n = 20) or HNSCC (n = 18) (Figure 3). Common metabolites that were identified in all ASCC sites were amino acids, fatty acids, carbohydrate, nitrogen compounds, and organic acids. OSCC and HNSCC appear to demonstrate similar metabolic profiles compared to LSCC (see Figures 3 and 4). Metabolic pathways commonly deregulated in both OSCC and HNSCC mainly concerned amino acid mobilisation, uptake, and polymerisation; lipid synthesis; and alternative energy. All twelve studies demonstrating BCAA deregulation were either OSCC or HNSCC. These compounds were increased in tissue and saliva but decreased in cancer patients’ plasma. Breakdown products of BCAA, alpha-ketoisocaprylic acid (KIC), alpha-ketoisovaleric acid (KIV), and alpha-ketomethylvaleric acid (KMV) were reported in three metabolomic studies [35, 42, 48] related to OSCC and HNSCC. However, the design and reporting heterogeneity meant these site-specific results should be approached with caution, and further detailed analyses were not performed.

3.5. Influence of Biosample Type on Metabolic Themes. There was an overall positive deflection in the proportion of metabolites present in the tissue, saliva, urine, and breath of ASCC patients and a negative deflection in plasma (Figure 3). This was particularly evident for amino acids. Metabolites in ASCC saliva samples were more similar to tissue than plasma. In particular, increased BCAAs were identified in tissue and saliva of patients with ASCC. In contrast, plasma BCAAs were decreased in ASCC plasma and not identified in urine or breath. This trend was also noted in other amino acids. BCKAs were decreased in the plasma of ASCC patients, but the proportions of fatty acids increase and decrease were similar in tissue samples. However, the design and reporting heterogeneity meant these sample-specific results should be approached with caution, and further detailed analyses were not performed.

4. Discussion

This systematic review provides an overview of progress in ASCC metabolic biomarker studies. The principal findings of this review were (i) favourable diagnostic performance of metabolic biomarkers for the detection of OSCC and HNSCC but not LSCC in pooled analysis, (ii) shared metabolic features of OSCC and HNSCC, and (iii) suggestion of a consistent role of the KEGG amino acid metabolic pathway in ASCC. Additionally, comparing sample types suggests metabolites are often depleted in the circulation and enriched in both tumour tissue and luminal surrogates, suggesting a model for ASCC biomarker kinetics. From the design perspective, clinical methodology and reporting quality was of a reasonable standard, but analytical methodology and reporting quality were often of a poor standard, and no studies performed exceptionally in both aspects.

ASCCs all have high mortality due to late disease detection. Currently, there are no screening strategies for any subtype of sufficient accuracy and quality to support political endorsement. Pooled analysis of identified studies regarding the detection of ASCC gave an area under the curve (AUC) of 0.927 with sensitivity of 85.7% (95% CI 78.9–92%), respectively. This diagnostic performance compares favourably to existing screening programmes such as faecal occult blood testing for colorectal cancer and cytological cervical screening test that currently are associated with lower sensitivity and specificity [61, 62]. Although the studies were generally...
of an exploratory nature without extensive validation, these results are encouraging and suggest metabolic biomarkers of ASCC may provide novel screening tools to identify high-risk populations, provided these efforts can progress to high-quality validation studies with appropriate power. The finding that the six studies reporting the highest AUC values had good clinical design and used targeted metabolomic methods suggests methodological rigour and hypothesis-driven metabolomics generate the best results.

In both discriminatory performance and metabolic themes, HNSCC and OSCC clustered away from LSCC, suggesting the underlying biology of those cancers is better suited to metabolic biomarker studies. Both HNSCC and OSCC arise from normally resident squamous cells, whereas LSCC arise from metaplastic squamous cells, perhaps explaining LSCCs’ relative metabolic heterogeneity. Moreover, genomic studies suggest LSCC to be distinct from HNSCC and OSCC [63–65]. Nonetheless, the relatively lower number of quantified metabolites for LSCC suggests this cancer warrants further study, perhaps using the similar analytical approaches from the best OSCC/HNSCC studies.

Despite using weighting to account for multiplicity from untargeted approaches, the heterogeneity in study design and quality, and the lack of independent validation, made comprehensive biological interpretation of the observed metabolic difference speculative. An exception to this was BCAA
metabolism, which was a consistent theme in across the ASCC subtypes. There were 36 significant differences in BCAA concentrations or their downstream metabolites, across 12 studies [25, 26, 28, 34–36, 41–43, 46, 56, 66]. This was far more than any other metabolite group. Increased BCAAs were observed in ASCC tissue samples from four studies [26, 36, 45, 53], and decreased BCAA levels in cancer-blood samples were observed in two studies [42, 56]. These differences were often stark. This suggests uptake of BCAAs into ASCCs against the concentration gradient. BCAAs constitute 35-40% of human protein and are thus essential amino acids necessary for protein synthesis in rapidly dividing cells [67]. They also have additional proproliferative effects. For example, leucine potently activates the mammalian target of rapamycin complex 1 [60, 68–70] and BCAA deamination is a major source of glutamine for alternative energy [68, 71]. Thus, BCAA metabolism is emerging as critical mediators of transformation and treatment escape in a number of malignancies including other squamous cancers [72, 73] and the present finding of consistent BCAA reprogramming in ASCC warrants further targeted study.

Metabolomic biomarker analytics has evolved considerably in the last 15 years, and that progress is reflected in the design heterogeneity of the included studies. Critical appraisal of analytical design using CAWG-MSI generally revealed a low standard. In contrast, only six of the studies included in this review demonstrated poor STARD/QUADAS-2 scores (score of less than 25), indicating a reasonable quality of clinical design and reporting (Supplementary Table 1). A key issue with metabolomic studies is the compromise between metabolomic coverage and unambiguous compound identification. Several studies used untargeted methodologies [16, 25, 26, 28, 29, 31, 33–38, 40–43, 45, 47–49, 51, 53] or more than one platform [3, 8, 9, 12] to increase their metabolomic coverage, although none used ultra-high coverage techniques such as Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR). No studies did not meet their objectives, or overstated their conclusions; however, this suggests that significant aspects of the ASCC metabolome have not been explored. Six studies that performed targeted methodology achieved level 1 identification of compounds of interest [27, 30, 32, 39, 46, 52], and two further studies used only commercially available spectral libraries for confirmation of their compound of interest [24, 56].

Additionally, our critical review has highlighted the following recurrent shortcomings in the current ASCC metabolomic literature: (i) lack of a clear sample size calculation; (ii) poor description of patient recruitment and inadequate description of clinical metadata; (iii) poor description of method validation; (iv) inconsistent quality assurance, especially replicate analysis; (v) biomarker performance frequently reported as multivariable models rather than clinical metrics; and (vi) lack of model validation data, either using internal cross-validation, or independent validation cohorts or studies. Using the CAWG-MSI checklist during study design would help to mitigate these issues [14].

A potential limitation was that more patients included in this review had late-stage disease (n = 331), and that the case mix was usually just reported rather than subject to subgroup analysis. Typically, the clinical motivation for the work was early cancer detection, which seems at odds with test populations enriched for late-stage disease, without subgroup analysis. However, the majority of these studies were performed in tertiary centre settings, which meant that patients would typically have been on a curative pathway. Thus, the observed metabolic differences can detect treatable disease, which provides a platform for further studies powered to detect truly early disease. It should also be noted that more than half of the articles in this review were performed in China and Japan and may not be applicable to Western populations.

5. Conclusion

This review summarised progress in using metabolites to identify patients with ASCC. There was significant heterogeneity in methodology and quality; however, especially for OSCC and HNSCC, metabolites showed promise for minimally invasive diagnosis. These two ASCC subtypes had similar metabolic phenotypes, with deregulation of amino acid metabolism particularly pronounced. Comparative analysis of different sample types suggested a kinetics model for amino acids across the endolumen. To aid the development of future studies and ultimately clinical translation, the summarised recurrent methodological weaknesses must be addressed, especially with respect to analytical design.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This project is funded by the Medical Research Council, UK. The funding reference number is MR/S022112/1.

Supplementary Materials

Supplementary Information Supplementary Figure 1: proportion of identified compound classes in different biosamples. Supplementary Table 1: summary of quality assessments. Supplementary Table 2: quality assessment of metabolic metadata based on CAWG-MSI guidelines. Supplementary Table 3: list of all metabolites, chemical class, and studies that identified them to be increased or decreased in biosamples. Supplementary Table 4: summary of significantly different metabolites listed in Supplementary Table 3. (Supplementary Materials)

References


[47] V. D. Kekatpure, N. Bs, H. Wang et al., "Elevated levels of urinary PGE-M are found in tobacco users and indicate a poor prognosis for oral squamous cell carcinoma patients," *Cancer Prevention Research*, vol. 9, no. 6, pp. 428–436, 2016.


