

# Silicon Nanowire Sensors Enable Diagnosis of Patients *via* Exhaled Breath

Nisreen Shehada<sup>(1)</sup>, John C. Cancilla<sup>(2)</sup>, Jose S. Torrecilla<sup>(2)</sup>, Enrique S. Pariente<sup>(2)</sup>,  
Gerald Brönstrup<sup>(3)</sup>, Silke Christiansen<sup>(3)</sup>, Douglas W. Johnson<sup>(4)</sup>, Marcis Leja<sup>(5-7)</sup>,  
Michael P. A. Davies<sup>(8)</sup>, Ori Liran<sup>(9)</sup>, Nir Peled<sup>(9)</sup>, and Hossam Haick<sup>\*(1)</sup>

<sup>(1)</sup> *Department of Chemical Engineering and Russell Berrie Nanotechnology Institute, Technion – Israel Institute of Technology, Haifa 3200003, Israel.*

<sup>(2)</sup> *Department of Chemical Engineering, Complutense University of Madrid, Madrid 28040, Spain.*

<sup>(3)</sup> *Max-Planck-Institute for the Science of Light, Günther-Scharowsky-Strasse 1, Erlangen 91058, Germany.*

<sup>(4)</sup> *Florida Radiation Oncology Group, Department of Radiation Oncology, Baptist Cancer Institute (BCI), 1235 San Marco Boulevard., Suite 100, Jacksonville, FL 32207, US.*

<sup>(5)</sup> *Faculty of Medicine, University of Latvia, 19 Raiņa boulv., LV1586, Riga, Latvia*

<sup>(6)</sup> *Department of Research, Riga East University Hospital, 6 Linezera iela, Riga, Latvia, LV1006*

<sup>(7)</sup> *Digestive Diseases Centre GASTRO, Riga, Latvia. 6 Linezera iela, Riga, Latvia, LV1006*

<sup>(8)</sup> *Molecular & Clinical Cancer Medicine, University of Liverpool, Apex Building, 6 West Derby Street, Liverpool, L7 8TX, United Kingdom.*

<sup>(9)</sup> *Thoracic Cancer Research and Detection Center, Sheba Medical Center, Tel Hashomer and Tel-Aviv University, Tel Aviv, Israel.*

**Keywords:** nanowire; sensor; disease; cancer; diagnosis; breath; volatile organic compound

**Abstract:**

Two of the biggest challenges in medicine today are the need to detect diseases in a non-invasive manner, and to differentiate between patients using a single diagnostic tool. The current study targets these two challenges by developing a molecularly-modified Silicon Nanowire Field Effect Transistors (SiNW FETs) and showing its use in the detection and classification of many disease breathprints (lung cancer, gastric cancer, asthma and Chronic Obstructive Pulmonary Disease). The fabricated SiNW FETs are characterized and optimized based on a training set that correlated their sensitivity and selectivity towards volatile organic compounds (VOCs) linked with diseased states. The best sensors obtained in the training set are then examined under real-world clinical conditions, using breath samples from 374 subjects. Analysis of the clinical samples showed that the optimized SiNW FETs can detect and discriminate between almost all binary comparisons of the diseases under examination with >80% accuracy. Overall, this approach has the potential to support detection of many diseases in a direct positive way, which can reassure patients and prevent numerous negative investigations.

Physicians are always challenged by the need to give the correct diagnosis as early in the onset of a disease is possible, whether the disease-related symptoms are absent or not evident.<sup>1</sup> Symptoms are not always characteristic of one particular disease; overlap of symptoms is common in, for example, lung diseases.<sup>2</sup> Patients with different respiratory diseases, such as malignant or benign tumors, or substantially less severe diseases, may have similar symptoms, e.g. cough, chest pain, difficulty to breathe, etc. These symptoms may be characteristic of lung cancer (LC), pneumonia, asthma, and chronic obstructive pulmonary disease (COPD).<sup>1,2</sup> Therefore, it is of particular clinical importance to find a diagnostic tool capable of distinguishing between these diseases. A diagnostic tool that involves no needle, surgery and/or active materials and/or radioactive exposure would have a benefit.

A highly promising approach that could meet the aforementioned need is based on the detection and classification of the disease breathprint, viz. the chemical profiles of highly- and semi-VOCs in exhaled breath linked with disease.<sup>3-15</sup> The rationale behind this approach relies on the fact that VOCs generated by cellular metabolic pathways during a specific disease circulate in the blood stream and diffuse into exhaled breath, which is easily sampled.<sup>4,16,17</sup> In certain instances, analysis of breathprints offers several potential advantages, such as: (a) breath samples are non-invasive and easy to obtain; (b) breath contains less complicated mixtures than either serum or urine; and (c) breath testing has the potential for direct and real-time diagnosis and monitoring.<sup>3,18-21</sup>

Several mass-spectrometry and spectroscopy studies have shown that the breathprint of a specific disease differs from that of healthy controls.<sup>22-26</sup> Some of the investigated VOCs are present in different concentrations in the breath of patients with a specific disease than in the breath of controls.<sup>22-26</sup> Spectrometry and spectroscopy techniques are powerful tools for detecting VOC of breathprints. However, to date, these techniques has been impeded by the need for expensive equipment, high levels of expertise to operate such instruments, the speed required for sampling and analysis, and the need for pre-concentration techniques.<sup>22,27-29</sup> For breathprint testing to become a clinical reality, several advances in the sensor development are needed. Chemical sensor matrices, based on nanomaterials, are more likely to become clinical and laboratory diagnostic tools, because they are significantly smaller, easier-to-use, and less expensive.<sup>18,30,31</sup> An ideal chemical sensor for breathprint analysis should be sensitive at very low VOC concentrations in the presence of water vapor because headspace of clinical samples is fully humidified. Furthermore, it should respond rapidly and differently to small changes in concentration, and provide a consistent output specific to a given exposure.<sup>18,30,31</sup>

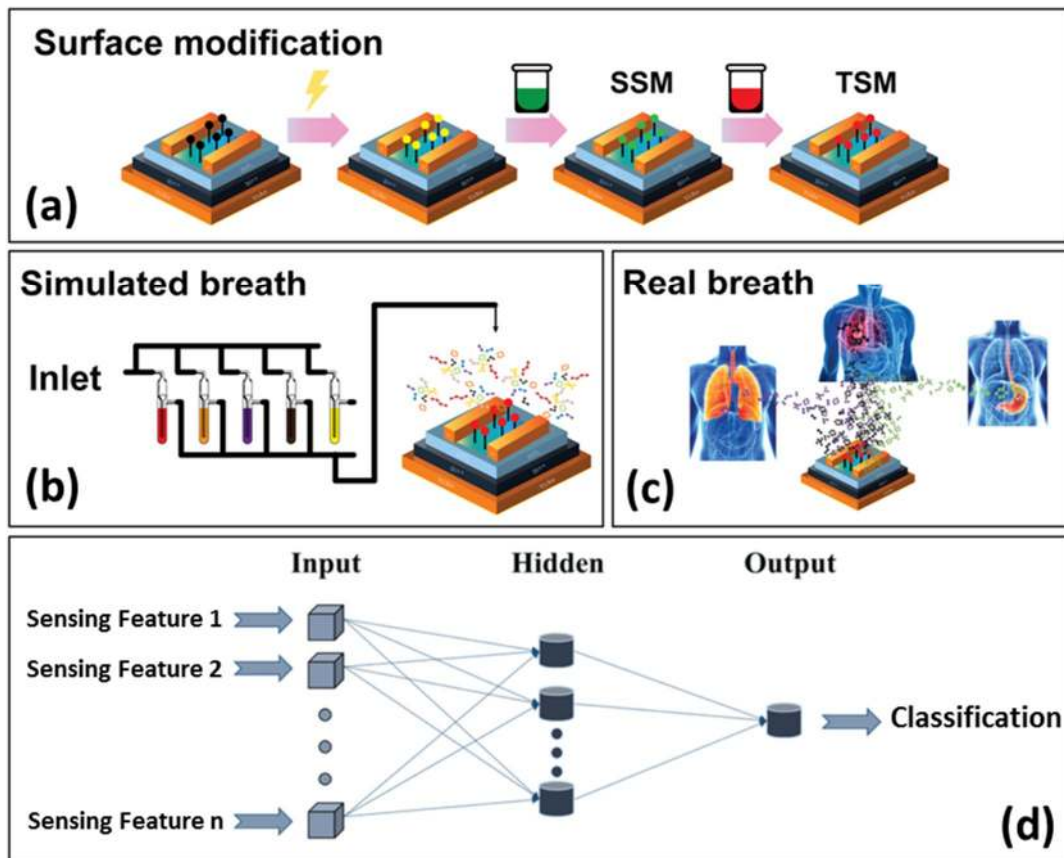
When not in contact with the analyte, the sensor should return to its baseline state rapidly, or be simple and inexpensive so that one could manufacture large numbers of disposable units.

We have developed Silicon Nanowire Field Effect Transistors (SiNW FETs)<sup>32-44</sup> as the sensing matrices for the detection and discrimination between disease breathprints. As representative diseases, we chose gastric cancer (**GC**), lung cancer (**LC**), asthma and Chronic Obstructive Pulmonary Disease (**COPD**). **NOTE:** For this study, the asthma and COPD are considered as a control group for LC, and is abbreviated as “**AC**” henceforth. These diseases cover both the direct and quasi-direct track to the breath.<sup>4,16</sup> Indeed, GC-related VOCs might reach the breath directly through the esophagus or from the lung alveoli. While LC- and AC-related VOCs reach the breath either by release directly into the airways or by diffusion in alveoli. The SiNW FETs are characterized and optimized based on a training set that correlated their sensitivity and selectivity towards the VOCs linked with the disease states under examination. The best sensors obtained in the training set are examined under real-world clinical conditions, using breath samples from 374 subjects. The high sensitivity, low-power consumption, fast-response times, and the compatibility with conventional silicon technology and readout circuitry, have the potential to get SiNW FETs to give us simple signal transductions of disease breathprints, as well as being amenable to miniaturization and scalability.<sup>45-50</sup>

## **Results and Discussion**

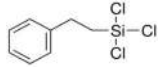
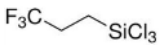
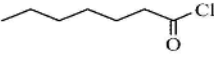
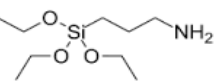
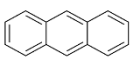
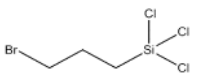
### **Preparation of molecularly-modified SiNW FETs**

The study consisted of 3 phases (**Figure 1a**). In the first, FETs based on well-aligned array of SiNWs (density  $\sim 1$  NW/  $100 \mu\text{m}^2$ ;  $40 \pm 8$  nm in diameter; and  $8.5 \pm 1.5 \mu\text{m}$  in length) were fabricated and characterized. Each SiNW FET was coated with different molecular modification (**Table 1**). The aim of the molecular modification was to passivate the surface states on the SiNW and optimize the interaction of VOCs with the SiNWs. Four SiNW FETs were coated with molecules having mostly non-polar (functional) side-groups (S1-S3 & S5) to improve their interaction with the non-polar VOCs in the breathprint. Two SiNW FETs were coated with molecules having mostly polar side groups (S4 & S6) to improve their interaction with the polar VOCs found in the breathprint.



**Figure 1.** Schematic representation of the experimental procedure divided into: **(a)** surface modification of the SiNW FET sensors (SSM: single-step modification; TSM: 2-step modification); **(b)** exposure of sensors to breathprint simulants, representative VOCs recognized as potential biomarkers of each disease; and **(c)** exposure to real breath samples of patients suffering from known diseases, compared with healthy control volunteers. **(d)** Representation of the artificial neural networks (ANN)<sup>33,51</sup> analysis method, where the sensing features are the inputs of the model, and the sample classification label is the output. ANNs are machine-learning algorithms that use a set of input vectors (sensing features in **Figure 1b**) to determine weighted parameters and mathematical functions, and to classify the samples (the sample classification label is the output).

**Table 1:** List of the molecular modifications used to coat the SiNWs - to optimize the interaction of the SiNW FETs with the breathprint-related VOCs

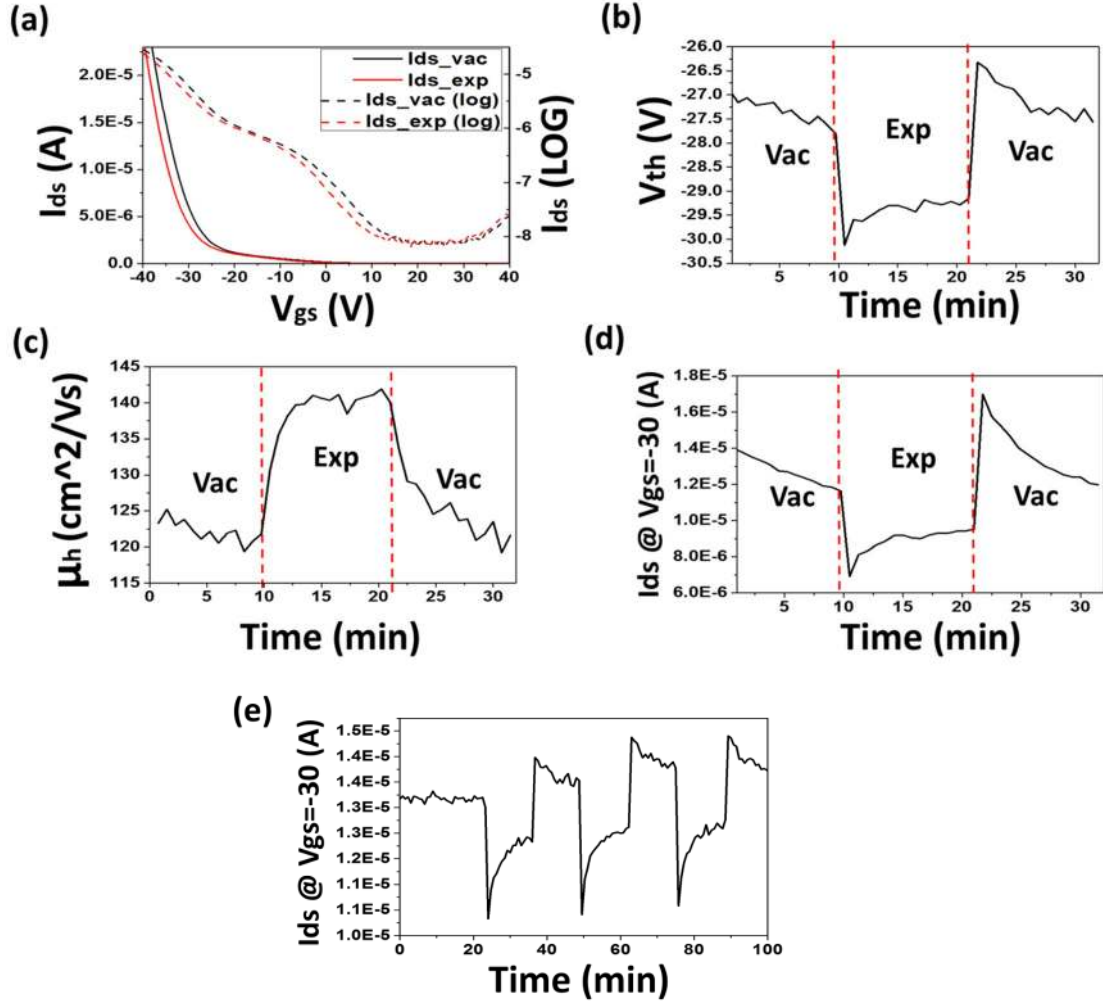
Sensor #	Modification	Structure
S1	Trichloro phenethyl silane (TPS)	
S2	Trichloro trifluoro propyl silane (TTPS)	
S3	Heptanoyl Chloride	
S4	3-Aminopropyl triethoxysilane (APTES)	
S5	Anthracene	
S6	Bromopropyl trichlorosilane (BPTS)	

### Testing the sensors exposed to simulants of disease breathprints

In the second phase of the study, the molecularly-modified SiNW FET sensors were exposed to a number of VOCs presumed as breathprint biomarkers for the chosen diseases: GC, LC and AC (**Table 2**). The presumed VOCs linked with GC conditions *via* breathprint are: 2-propenenitrile, furfural, 6-methyl-5-hepten-2-one.<sup>14,31</sup> The presumed VOC linked with AC conditions by breathprint is pentane.<sup>31</sup> The presumed VOCs linked with LC conditions by breathprint are: heptane, decane, 2-methylpentane, 2-ethyl-1-hexanol, propanal, pentanal and acetone.<sup>31</sup> Each SiNW FET was exposed to 4 concentrations of each VOC, and each exposure was repeated 3 times. In each exposure to VOC, the source-drain current ( $I_{ds}$ ) vs. back gate voltage ( $V_{gs}$ ) characteristic curve was obtained by sweeping the gate voltage between +40V and -40V (**Figure 2a**). For the sake of reference,  $I_{ds}$ - $V_{gs}$  curves were obtained in vacuum before and after exposure to the VOC (**Figure 2a**).

**Table 2:** List of the representative VOCs linked with gastric cancer (GC), asthma & COPD (AC), or lung cancer (LC) from breathprint showing significant statistical differences ( $p < 0.05$  in previous clinical studies). The VOC concentrations exposed to the molecularly-modified SiNW FETs, as well as the transition from concentration in  $p_a/p_o$  terms ( $p_a$  is the partial pressure of the VOC and  $p_o$  is the vapor pressure of the VOC) to parts per billion units (ppb), are presented.

Disease	VOC	$P_a$ (mmHg)	Concentrations tested ( $p_a/p_o$ )	Concentrations tested (ppb)
GC	2-Propenenitril	83	0.001-0.008	100-870
	Furfural	2	0.001-0.008	2.6-21
	6-Methyl-5-heptene-2-one	0.8	0.001-0.008	0.98-7.9
AC	Pentane	6.2	0.001-0.008	8-65
LC	Heptane	40	0.001-0.008	50-420
	Decane	1.5	0.001-0.008	2-15
	2-Methylpentane	350.3	0.001-0.008	460-3680
	2-Ethyl-1-hexanol	0.23	0.001-0.008	0.29-2.36
	Propanal	258	0.0005-0.004	170-1360
	Pentanal	26	0.001-0.008	34-273
	Acetone	230	0.001-0.008	300-2420



**Figure 2.** (a) Characteristic  $I_{ds}$ - $V_{gs}$  curve of SiNW FET modified with heptanoyl chloride (S3) in vacuum (black line) and on exposure to 2-propenenitrile at a concentration of  $p_a/p_0=0.001$  (red line). Curves are presented in both linear (solid line) and logarithmic (dashed line) scales. Sensing response of: (b)  $V_{th}$ , (c)  $\mu_h$  and (d)  $I_{ds}$  @  $V_{gs} = -30V$  on exposure to 2-propenenitrile at a concentration of  $p_a/p_0=0.001$ . “Vac” stands for “vacuum” and “Exp” stands for exposure. (e) 3 repetitive response cycles of  $I_{ds}$  @  $V_{gs} = -30V$  exposed to 2-propenenitrile ( $p_a/p_0=0.001$ ) condition for evaluating the repeatability of the sensing signals.

In most instances, exposing the SiNW FETs to VOCs resulted in a change in the  $I_{ds}$ - $V_{gs}$  curves. **Figure 2a** presents the characteristic  $I_{ds}$ - $V_{gs}$  curve of a SiNW FET modified with heptanoyl chloride (S3) under vacuum (black lines) and under exposure to  $p_a/p_0=0.001$  of 2-propenenitrile (red lines). The FET curve changes on exposure to 2-propenenitrile. Closer examination of the curves shows that several FET features are changed on exposure, which include the voltage threshold ( $V_{th}$ ; the voltage under which the charge transfer regime changes from depletion to accumulation), the charge carrier (hole) mobility ( $\mu_h$ ; the velocity of the charge



carriers moving under the influence of the electrical field), and the  $I_{ds}$  @ different  $V_{gs}$  values. The extraction and calculation of these features is described in the Experimental Methods section. As these features change with exposure, their values (which can be monitored and recorded during the experiment) may be informative. For example, the  $V_{th}$  values decrease on exposure to 2-propenenitrile from -27 to -29.5V (**Figure 2b**) at the time the  $\mu_h$  values increase from 125 to 140  $\text{cm}^2/\text{V}\cdot\text{s}$  (**Figure 2c**), and the  $I_{ds}$  @  $V_{gs}=-30\text{V}$  values decrease slightly from 120 to 80  $\mu\text{A}$  (**Figure 2d**). In the representative example (**Figure 2e**), repeated exposure cycles on the same sensor are usually characterized by a relatively stable baseline and an almost repeatable sensing features from cycle to cycle ( $\pm 3\%$  variance).

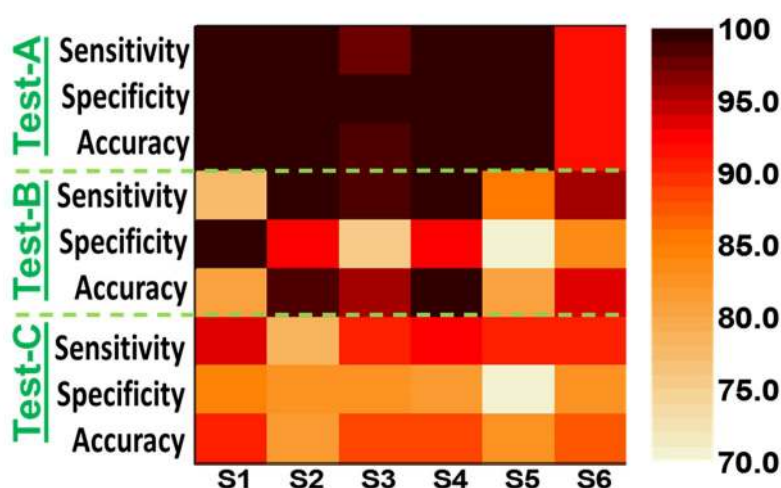
The detection limit of the sensors to each VOC as well as the concentrations found in breath can be found in the Supporting Information (**Table S1**). The detection limit of each sensor is different and depends on the VOC's structure. In this context, it is important to clarify that the SiNW FET sensors do not obey the lock-and-key sensing approach.<sup>18,31</sup> Instead, they have affinity to multiple VOCs and, therefore, the sensing signal on exposure to a mixture of VOCs (e.g., a breath sample) reflects the fingerprint of *all* the VOCs in it. Under these circumstances, different molecular modifications of the SiNW change the affinity balance between the different VOCs found in a specific mixture. Therefore, even if the detection limit of specific SiNW FET to a certain VOC is higher than its concentration in the breath, the SiNW FET can still detect the change in breath sample composition, as the overall change (sum) of VOC changes in the breath due to disease state is in the range of a few to 10s of ppm. These SiNW FETs have a very wide dynamic range between their limit of detection (Table 1) up to thousands of ppm, when a thin film of condensed liquid begins to develop on the surface of the sensors.

The ability to separate disease-related VOCs by a single sensor stems from their ability to respond simultaneously to different VOCs, each of which induces a different effect on the FET features due to one or a combination of the following scenarios: **(i)** charge-charge interactions between the functional group of the modified SiNW, which changes the charge density on the surface of the SiNW, resulting in a change to  $V_{th}$ ,  $\mu_h$  or the currents in the linear region; **(ii)** addition of polar or polarizable molecules on the surface of the SiNW, which changes the electrical field on the surface, resulting in a change to the value of the subthreshold currents and other features, similarly to molecular gating phenomena previously reported;<sup>35</sup> **(iii)** passivation of the surface states by the VOCs.<sup>34,35</sup>

Investigation of the change in the entire  $I_{ds}-V_{gs}$ , mainly the depletion and accumulation features, could help significantly an increase in the number of multiple

sensing features. These features operate as “virtual sensors”, viz. a set of independent sensing signals extracted from the electrical measurements of the device, and can be monitored separately as if they are individual sensors. This behavior leads to a decrease in the number of sensors needed for classification, and thus simplifies the system in terms of its supporting circuitry and infrastructure. To avoid unnecessary computation, a process of feature selection was applied by choosing only features with a signal-to-noise ratio >3 and by choosing only features that represent physical phenomena in the device.

Following the initial feature selection, artificial neural network (ANN) analysis was applied. Its purpose was to test the ability of the sensors to correlate VOCs with the disease to which they should relate, and to assess which sensor would be more suitable for disease diagnosis. ANN was applied on the basis of 36 samples for GC simulation, 12 samples for AC simulation, and 84 samples for LC simulation. Separation of GC-related VOCs from AC biomarkers was called “**Test-A**”; separation of LC-related VOCs from AC-related VOCs was **Test-B**”; and separation between LC- and GC-related VOCs was “**Test-C**”. In **Test-A** and **Test-B**, the cancer group was considered the “positive” group, whereas AC was considered the “negative” group. In **Test-C**, LC was considered the “positive” group and GC was considered “negative”. The success in classification of the ANN analysis in each test is presented in **Figure 3**.



**Figure 3.** Classification results of sensing signals obtained from the molecularly-modified SiNW FETs upon exposure to a simulated VOC-related breathprint of the diseases. The color bar indicates the percentage of correct classification of samples comparing VOC of GC vs AC (Test-A); LC vs AC (Test-B); and LC vs GC (Test-C).

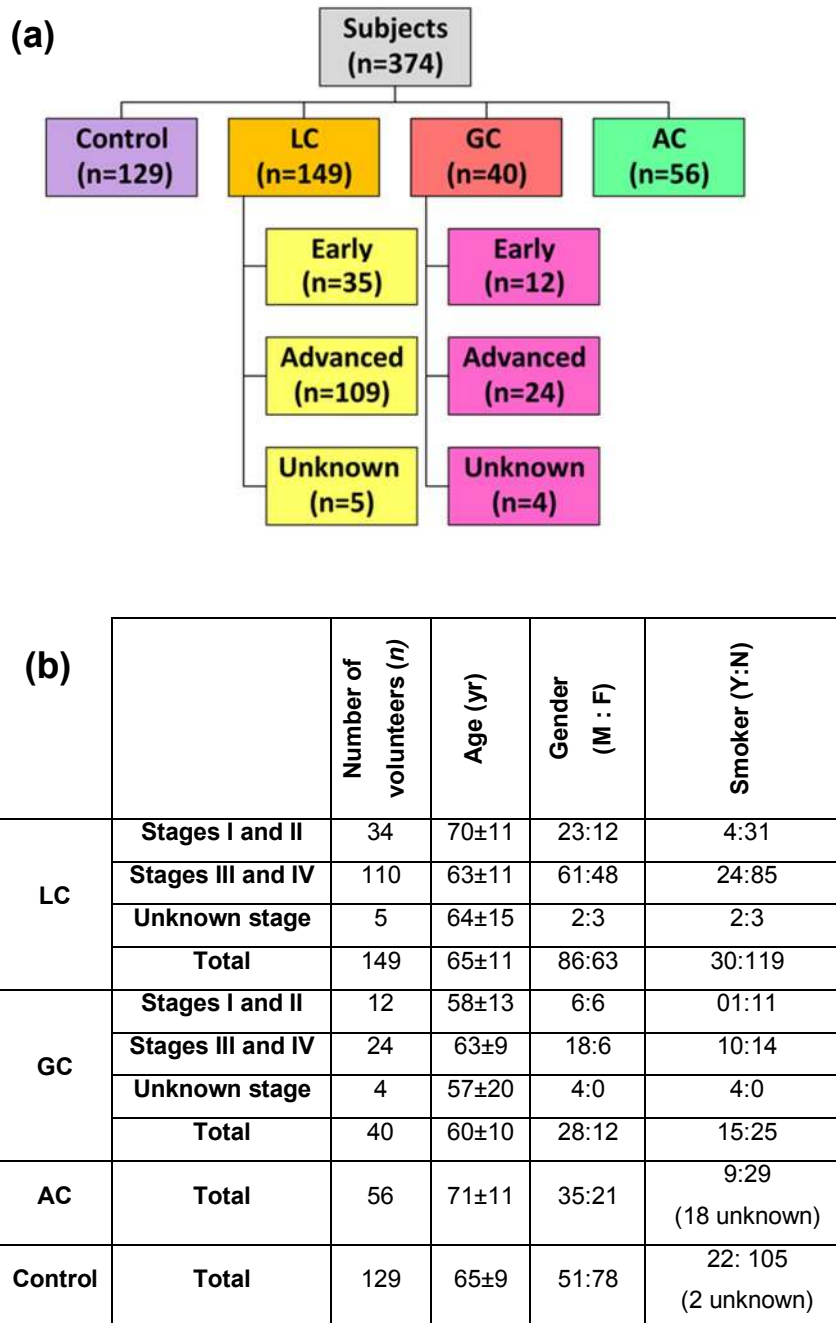
The ANN model calculated values of sensitivity, specificity and accuracy of each comparison, the method being found in the Experimental Methods section. Results from the vast majority of the sensors (**Figure 3**), and in most comparisons, were satisfactory (>80% accuracy). Besides this accuracy for the classification, a few important findings were recorded. In **Test-A**, most sensors correctly classified 100% of the cases, irrespective of the vapor pressure values of the VOCs. In **Test-B**, S4 had a similar classification ability to S2 and S3 for LC (~100% accuracy) and gave slightly better classification than S1, S5 and S6. On the other hand, S1 showed the highest classification ability of AC samples (100%). In **Test-C**, S1 again gave the highest percentage of correct classification; 93% for LC-related VOCs and 85% for GC-related VOCs, resulting in 92% accuracy.

### **Validation of the sensors in a clinical study**

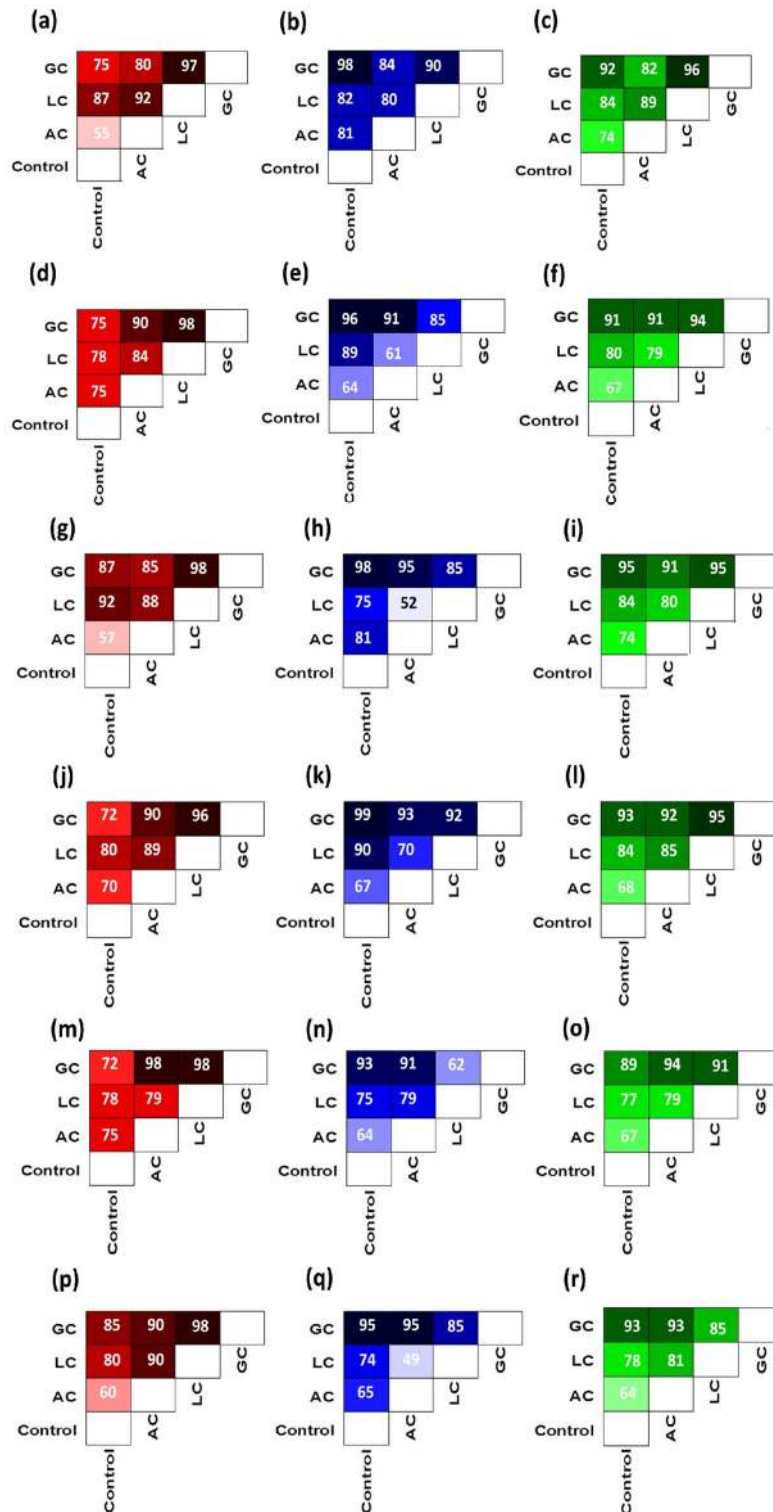
To validate the performance of the SiNW FETs under lab conditions using simulants of breathprint-related VOCs, clinical trials were carried out with the same sensors (Experimental Methods section); the design of the clinical study is shown in **Figure 4**. Briefly, breath samples were collected from 374 volunteers belonging to 4 groups: control subjects not suffering from any of the conditions tested (n=129), LC patients (n=149), GC patients (n=40) and AC patients (n=56). Cancer patients were also divided into 2 groups according to the staging of the cancer: early (stages 1 and 2) and advanced (stages 3 and 4). Additional information was also collected from the volunteers, including age, gender, smoking status, *etc.*, shown in **Figure 4b**. Clinical studies provides 3 main advantages over the simulated breathprint studies, because they can precisely determine: **(i)** the signature of all VOCs of the disease breathprint on the SiNW FET sensor; **(ii)** correlation between the sensor's sensitivity and specificity of the disease breathprint to its existence in a population of patients; and **(iii)** the necessary feedback that accounts for confounding factors, such patients' diet, metabolic state, genetics, *etc.*

For the clinical study, 2 kinds analysis were carried out: ANN analysis and discriminant function analysis (DFA). As mentioned earlier, ANNs are a machine-learning method inspired by biological neural networks (*i.e.* the human nervous system). It is based on a set of functions connecting the input (sensor features in this case) with the output (classification of samples to a specific disease). On the other hand, DFA is a linear supervised pattern recognition method that effectively reduces the multidimensional experimental data, in which the classes to be discriminated are defined, before carrying out the analysis. More details on these approaches and their implementation in the clinical study can be found in the Experimental Methods

section. Using one or both of these approaches, the aim was to distinguish between breathprints of patients belonging to the 4 study groups (**Figure 4**).



**Figure 4.** (a) Division of subjects into the groups. The 374 subjects who provided breath samples were separated into the following groups: healthy control volunteers (n=129), LC patients (n=149), GC (n=40) patients and AC patients (n=56). Cancer patients were further separated into 2 groups according to the disease stage: Early and advanced. (b) Full clinical details including age, gender, smoking status and as r classification into one of the groups given in (a).



**Figure 5.** Sensitivity (red; a, d, g, j, m, p), specificity (blue; b, e, h, k, n, q) and accuracy (green; c, f, i, l, o, r) for binary comparisons between the response of S1-S6 to breath samples of volunteers from all groups: LC, GC, AC and healthy control. Classification was obtained by ANN analysis (each row corresponds to a different sensor- top row (a, b, c) corresponds to S1, the second (d, e, f) to S2, etc.). In this classification, all stages of a disease were considered as one group (without staging). Numbers in white are the values, e.g. the value in the red box relating GC and control is the sensitivity of their comparison; the

value in the blue box is the specificity and in the green box is the accuracy. **NOTE:** In comparing any of the groups with the "control" group, the latter was clearly considered the "negative" group, while the others were considered the "positive" groups. In comparing one of the cancer groups (LC or GC) to AC, former was considered the "positive" and AC the "negative" group. Finally, in comparing the 2 types of cancer, LC was taken as the "positive" group, and GC the "negative" group. The values of the comparisons are reviewed in Supporting Information **Table S3**.

**Figure 5** presents the sensitivity, specificity and accuracy values as determined by ANN analysis for binary comparisons between breathprint of volunteers from all groups. Most sensors had an acceptable ability to classify the different diseases, distinguishing them from the control samples. S1, S3 and S5 have the highest ability of to classify correctly breathprints. In trying to separate the LC from the control group, S1 was the most suitable sensor, with sensitivity, specificity and accuracy all >80% (sensitivity 87%, specificity 82% and accuracy 84%). In trying to separate GC patients from the control group, S3 had greater sensitivity than S1, sensitivity 87%, specificity 98%, and accuracy 95%. Regarding AC, when the goal was to rule in a subject (meaning the patient was most likely sick and the test being used as confirmation), the sensor of choice was S1 (specificity 81%). When the goal was to rule out a patient, the sensor of choice was S5 (sensitivity 75%). Regarding the separation of LC patients from AC patients (cancer vs. non-cancer), S1 discriminated best overall (sensitivity 92%; specificity 80%; and overall accuracy 89%) - making it an excellent sensor for ruling out diagnostic decision. Examining the discrimination between GC and AC, most sensors showed noteworthy results (accuracy >80%). Nevertheless, S5 gave the best results, with all 3 statistical parameters >91%, with an especially high sensitivity of 98%. Comparing GC and LC patients, S1 gave the best results, offering extremely high values of sensitivity, specificity and accuracy (97, 90, and 96%, respectively). Looking at the separation ability of all the sensors in the case of AC vs. control, the performance of the sensors was rather low across the board. This could have been to only one VOC (pentane) characterizing these diseases (asthma and COPD; cf. **Table 2**) rather than a pattern of VOC, as in the other diseases. Adsorption of a combination of VOCs on the surface of the Si NW sensor led to a higher signal-to-noise ratio.

The performance of our devices under real-world conditions has several advantages over the current gold standard test. For example, the National Lung Screening Trial found that CT scans were highly sensitive in detecting lung cancer in smokers compared with chest x-rays, but they were not very specific in ruling out

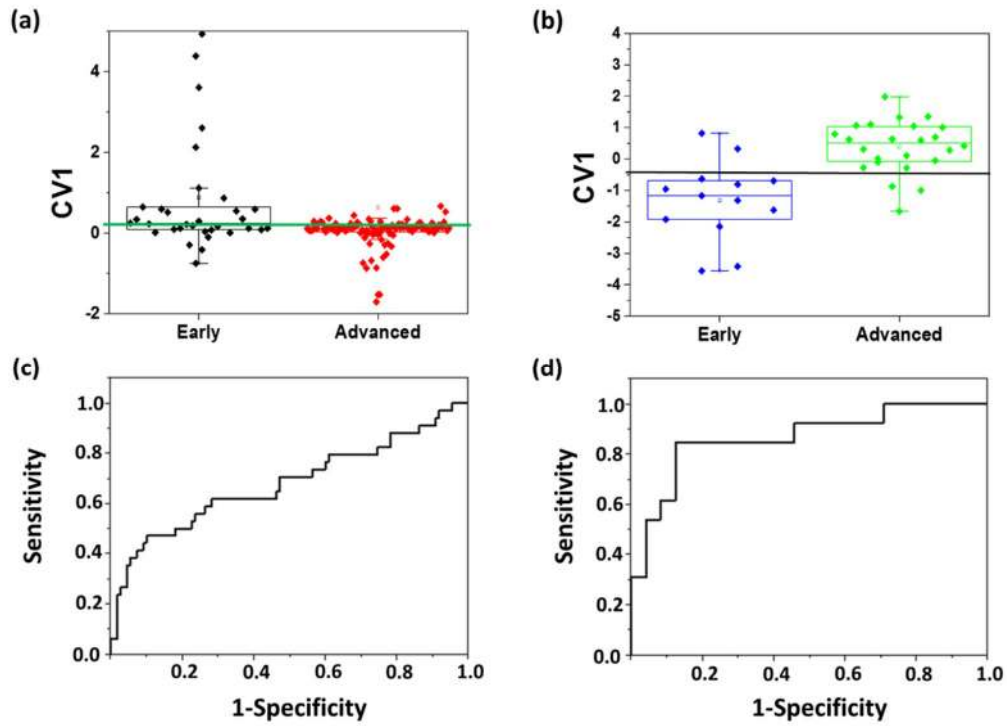
malignancy.<sup>52</sup> Sensitivity was 94% and specificity 73% for lung cancer detection with CT compared with 74% and 91% with chest x-rays in the first round of screening of high-risk smokers and former smokers included in the trial.<sup>52</sup> This combination led to substantially more positives in the CT group (27 vs. 9%), nearly all of which prompted follow-up diagnostic procedures.<sup>52</sup> In comparison, the diagnostic performance of S1 provided a benefit in being more cost efficient, practical, and efficacious. In another example, upper endoscopy with a full biopsy is the gold standard for diagnosing and screening gastric cancer and the related precancerous lesions.<sup>53</sup> However, this is not a feasible screening approach outside Asia and the cost-effectiveness in other parts of the world has not been assessed. The cost-effectiveness of our sensors would thus be more affordable and give greater compliance for an at-risk population. On the other hand, an acceptable non-invasive test with high diagnostic performances is lacking, since the best available test – detection of pepsinogens in blood – fails to reach the expected accuracy levels – it could be missing in the majority of cancer cases.<sup>53</sup> Whilst in population-based screening settings, the sensitivity of pepsinogen tests is 67-85% and the specificity 76-87% for the detection of atrophy, the sensitivity of gastric cancer detection is only 37-62%.<sup>53</sup> In comparison to these values, S3, for example, is the frontrunner for clinical use as population-based diagnostic or screening tool.

To further validate the analysis, the same binary comparisons were carried by discriminant function analysis (DFA). The data were similar to the ANN results, with S1 being the most adequate sensor for disease classification in most cases, coupled with S3 and S4 for the remaining classifications (LC vs. control and LC vs. AC). Results of the DFA analysis are given in Supporting Information, **Figure S1 and Table S4**. It is worthy to point out that looking at the DFA results (Figure S1), we get a different picture than received in using ANN. In the case of DFA, it is S3 that has the best results for discriminating between GC and the other two diseases, resulting in an accuracy of 95%. When comparing GC with the control group, we get an accuracy of 93%, which is very high, but S4 showed better results (accuracy of 94%) due to a higher sensitivity. When studying the differentiation of LC from the control group, we can see that a combination of S3 and S4 should be used, as S3 has a high value of specificity (87%), and S4 has a high sensitivity value (90%). Therefore, S3 should be used as a ruling in tool, while S4 would be used for ruling out. Separating the cancerous lung conditions from the non-cancerous lung conditions (AC), S4 is the most suitable sensor as it has extremely high values of sensitivity, specificity and accuracy (89% sensitivity, 75% specificity, and 86% accuracy). The last comparison to be made is the AC group and the control group. In this case, we can see low

sensitivity values all across the board, similarly to the values received when using ANN. Despite that fact, we still have satisfactory accuracy values, due to the high specificity. The most suitable sensor to be used in this case is either S3 or S4, as they both exhibited the highest specificity and accuracy values (91% and 78% accordingly).

Following the very successful classification of the breathprints in the clinical study, we used DFA to classify breath samples collected from cancer patients according to the stage of the disease. For this comparison, DFA was preferred to ANN because the latter does not provide reliable results with small size populations.<sup>54</sup> The best results were achieved in the DFA model built by using features extracted from S4 for both LC and GC staging (**Table 4**). During the development of the staging models, early stages were considered to be the positive group and the advanced stages negative, which means that the sensitivity indicates detection at the early stage, whereas the specificity is related to advanced stage detection. Classification of the results in the staging analysis by DFA are shown in **Figure 6**. In interpreting this data, DFA classification gave an accuracy of 81% for LC staging and 87% for GC staging. In LC staging, the sensitivity value was low (34.5%), most likely due to the difference in the size of the groups (n=34 for early stage patients and n=110 for the advanced stage). Indeed, big differences in the group size leads to the result being biased towards the larger group, leading to a lack of sensitivity.<sup>55</sup> Despite low sensitivity, the specificity achieved by S4 was very high (95%), being the highest accuracy value. Regarding the GC staging results, S4 showed an equally high ability in identifying both early and advanced stages of the disease (86.5% of the samples were correctly classified; 84.6% of the early stage; and 87.5% of the advanced stage). The AUC values of the classification achieved by S4 were calculated as 0.68 for LC staging and 0.87 for GC staging. ROC curves developed for the separation of each stage from the negative samples can be found in Supporting Information, **Figure S2**. Clear differences were seen between each set of ROC curves (early vs. negative compared to advanced vs. negative) and the different diseases.





**Figure 6.** Box plot of the first canonical variable (CV1) values in the staging of **(a)** LC and **(b)** GC, as calculated using the S4 DFA model. The box represents 95% confidence interval of CV values; error bars represent the standard deviation. The central lines represent Youden's cut-point. "Early" stages refer to stages 1 and 2 of the disease (localized tumors), and "advanced" are stages 3 and 4 (metastasized tumors). Receiver operating characteristic (ROC) curves **(c&d)** are also presented. The P-values of these discriminations were 0.0017 for **(c)** LC and **(d)** 0.033 for GC (d).

**Table 4:** Summary of sensitivity, specificity and accuracy values achieved from DFA analysis for the staging of lung cancer (LC) and gastric cancer (GC)

	<b>S1</b>			<b>S2</b>			<b>S3</b>			<b>S4</b>			<b>S5</b>			<b>S6</b>		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
<b>LC-Staging</b>	20.6	94.5	<b>77.1</b>	26.5	88.2	<b>73.6</b>	32.5	94.5	<b>79.9</b>	35.3	94.5	<b>80.6</b>	35.3	90.9	<b>77.8</b>	29.4	88.2	<b>74.3</b>
<b>GC- Staging</b>	53.8	83.3	<b>73</b>	61.5	83.3	<b>75.7</b>	77	80	<b>78.4</b>	84.6	87.5	<b>86.5</b>	53.8	83.3	<b>73</b>	53.9	87.5	<b>85.7</b>

### **Summary and Conclusions:**

Molecularly-modified SiNW FETs linked with pattern recognition methods and/or machine learning have successfully demonstrated the detection and classification of several disease breathprints from lung cancer, gastric cancer, asthma and Chronic Obstructive Pulmonary Disease patients, both under lab and clinical conditions. SiNW FET coated with trichloro(phenethyl)silane (TPS) (S1) and heptanoyl chloride (S3) could detect VOCs linked with the breathprint of the diseases and distinguish one from another. These sensors could separate early stages of cancer from advanced stages, allowing not only the detection of the disease, but estimation of its progress. This information can supply clinicians with valuable and useful information in diagnosis and treatment planning without the necessity of causing patients discomfort or pain, preventing time delay and reducing the financial cost incurred by the clinic/hospital. S1 and the others giving satisfying performances (e.g., S5) could serve as translational tools from fundamental research to point-of-care because the sensors are significantly smaller, easier-to-use, and less expensive than other sensing approaches. They could be used as point-of-care and wearable diagnostic devices. But this pilot study does not allow us to draw far-reaching conclusions. A multi-center clinical trial with a considerably increased sample size, using Si NW FET sensors that enable *in-situ* sampling and analysis, is required to confirm the breathprints. We believe the reported SiNW FET sensor can be modified for selective VOC recognition and concentration prediction in other cancer or disease states.

## **Experimental Methods**

### ***Growth of the SiNWs***

P-type silicon nanowires (SiNWs) were grown following the vapor liquid solid (VLS) principle using a cold-wall chemical vapor deposition (CVD) reactor. During the 30 min CVD process under a total pressure of 2 mbar, the flow rates of SiH<sub>2</sub>, B<sub>2</sub>H<sub>6</sub> (100 ppm in H<sub>2</sub>), and Ar were kept constant at 4, 1, and 10 sccm, respectively, *i.e.* in a B to Si ratio of 1:20,000. Growth was catalyzed by commercially available Au colloids (British Biocell Int.) of 30 nm diameter. Growth substrate temperature was ~500°C. The resulting SiNWs had an average diameter of 40±8nm and average length of 8.5±1.5 μm.

### ***Deposition of the SiNWs array***

The as-grown SiNWs were first immersed in a KI/I<sub>2</sub>/H<sub>2</sub>O solution (mass ratio 4:1:40) for 1 min to remove the gold catalyst used in the growing process, and any possible gold contaminants remaining on the surface of the SiNW. The Si NWs were etched using buffered hydrofluoric acid for 10 sec. Following this, the SiNWs were dispersed in ethanol using ultra-sonication for 5 sec, and were later deposited on a highly-doped (0.001 Ω-cm resistivity) P-type Silicon (100) substrate with 300 nm thermal oxide and a Ti/Au (10/200 nm) bottom gate electrode. Deposition was based on a spray-coating procedure,<sup>56</sup> which started with the deposition of the substrate on a hotplate set at 75°C. A SiNW suspension was applied with a spray gun (PRONA R2F) at a carrier gas pressure of 40 psi, and a tilt angle of (5°±2°) to the substrate. The spray-coating process resulted in well-aligned nanowire arrays (density ~ 1 NW/100 μm<sup>2</sup>).

### ***Fabrication of SiNW FETs***

The substrate coated with SiNWs was cleaned by successive immersions in acetone, methanol, and isopropanol, followed by 5 sec rinsing in buffered HF. The top source and drain electrodes (20 pairs of 20 nm Ti/ 160 nm Au interdigitated electrodes, length 1300 μm, width 2 μm, spacing 2 μm) were deposited by a photolithography (Karl Suss MA6 Mask aligner) and lift-off process on top of the SiNWs. Surface characterization of the fabricated used dark-field light microscopy to determine the density of SiNWs on the device, which were then observed in a scanning electron microscope (SEM) to assure proper contact between the SiNWs and the electrodes.

### **Surface modification of the SiNW FETs**

Surface modification of the SiNW FETs was carried as described before.<sup>43</sup> Briefly, following surface activation achieved by 30 sec of plasma treatment, the SiNW FET surfaces were modified with various molecules (listed in Table 1). SiNW FETs were modified using a single-step silane modification (SSM), a 2-step silane-chloride modification (TSCM), or a 2-step silane-monomer modification (TSM). Devices modified using the SSM approach were: (i) immersed in a silane/chloroform (silane/ethanol in case of APTES) solution (2 mM; 10 mL) for 45 min in room temperature, and then (ii) successively rinsed with chloroform, acetone, ethanol, isopropanol (ethanol, acetone, ethanol and isopropanol in case of APTES) and drying by N<sub>2</sub> flow. The silane molecules prepared by this SSM approach and used in this study were trichloro(3,3,3-trifluoropropyl)silane (CF<sub>3</sub>CH<sub>2</sub>CH<sub>2</sub>SiCl<sub>3</sub>; TTPS), trichloro(phenethyl)silane (C<sub>6</sub>H<sub>5</sub>CH<sub>2</sub>SiCl<sub>3</sub>; TPS), 3-aminopropyl-triethoxysilane (C<sub>9</sub>H<sub>23</sub>NO<sub>3</sub>Si; APTES) and 3-bromopropyl trichlorosilane (C<sub>3</sub>H<sub>6</sub>BrCl<sub>3</sub>Si; BPTS). SiNW FETs modified using the TSCM approach were: (i) immersed in 3-aminopropyl-triethoxysilane (APTES)/dehydrated ethanol (10 mM, 20 mL) for 1 h, (ii) rinsed with ethanol, acetone, isopropanol and dried by N<sub>2</sub> flow, (iii) immersed in a solution of acylchloride in chloroform (10 mM) with 10 μL of triethylamine for 17 h, and finally, (iv) successively rinsed with chloroform, acetone, ethanol, and isopropanol before being dried by N<sub>2</sub> flow. The molecular modification prepared by the TSCM approach was heptanoyl chloride (C<sub>7</sub>H<sub>13</sub>ClO). SiNW FETs modified using TSM were: (i) immersed in trichloro(3,3,3-trifluoropropyl)silane (CF<sub>3</sub>CH<sub>2</sub>CH<sub>2</sub>SiCl<sub>3</sub>; TTPS)/chloroform (2 mM, 10 mL) for 1 h, (ii) successively rinsed with chloroform, acetone, ethanol and isopropanol, before being dried by N<sub>2</sub> flow, (iii) drop-casted with a monomer solution in tetrahydrofuran (THF) (1 mM, 2 μL) and kept in a vacuum oven overnight (55°C), and (iv) rinsed with THF, acetone, ethanol and isopropanol before drying in a N<sub>2</sub> flow. The molecular modification prepared by the TSM approach was anthracene (C<sub>14</sub>H<sub>10</sub>).

### **Breath sample collection**

Exhaled alveolar breath samples from 374 volunteers were collected. The volunteers were divided into 4 groups: 149 volunteers with lung cancer; 40 with gastric cancer; 56 volunteers with non-cancerous lung diseases (asthma, COPD or both); and 129 volunteers who were negative for all these diseases. These samples were collected in 4 locations: (i) Riga East University Hospital, Riga, Latvia; (ii) Cancer Research Center, University of Liverpool, Liverpool, UK; (iii) Thoracic Cancer Research and Detection Center, Sheba Medical Center, Tel-Hashomer, Israel; and (iv) Baptist

Cancer Institute, Jacksonville, Florida, USA. A summary of the clinical data is presented (**Figure 4b**).

In each location, the samples were collected in the same hospital room. Patients were asked to fast, and withhold smoking and alcohol consumption for at least 2 h before sampling. The sorbent tubes were stored at 4°C until transported to Laboratory of Nanomaterial Based Devices (Technion – Israel Institute of Technology, Haifa, Israel) for analysis. The duration between sample collection and analysis was <3 months (NOTE: breath VOCs can be preserved in these sorbent tubes up to 6 months without change in their composition). The breath samples were collected following a previously described protocol.<sup>14,57</sup> To reduce the effect of ambient contamination, a lung washout procedure was used. According to the protocol, the patient took regular, unexerted breaths for 3 min through a mouthpiece with a filter cartridge on the inspiratory port (Eco Medics, Duerten, Switzerland), thereby greatly reducing the concentration of ambient VOCs in the inspired air. **NOTE:** The subjects were asked to breathe normally throughout the entire procedure to avoid hyperventilation. Following lung washout, the subjects were asked to inhale normally once more through the filter and exhale normally through the mouthpiece into a separate exhalation port against 10-15 cm H<sub>2</sub>O pressure (ensuring the closure of the vellum to exclude contamination through nasal entrainment). During exhalation, the respiratory dead space air is first exhaled, the alveolar air coming thereafter. The dead-space air was automatically filled into a designated bag at the beginning of the exhalation and was later removed. The alveolar breath from the tidal end was filled into breath collection bags (China: TedlarR bags, Keika Ventures, LLC; Latvia: Mylar bags, Quintron, Milwaukee, WI, USA). The breath collection method applied is a single-step process; the volunteer does not have to take care of exchanging the dead space bag with the collection bag. Following the collection procedures, the breath samples were transferred to ORBOTM 402 Tenax sorption tubes (specially treated; 35/60 mesh; 100/50 mg; from Sigma-Aldrich, USA) by pumping each sample through a sorbent tube. We have confirmed that the VOCs in the breath of lung cancer patients and other lung conditions can be trapped and stored in the ORBOTM 402 Tenax® TA sorption tubes.

#### ***Breath analysis with the SiNW FET sensors***

A manual thermal desorption (TD) system was used to transfer the VOCs trapped in the disposable sorbent tubes. Each tube was heated at 190°C for 8 min under a constant N<sub>2</sub> (99.999% purity) flow at 60 ml/min to release the trapped breath VOCs into the bag. A pulse of the content of the bags was delivered into a stainless steel

test chamber containing the SiNW FET sensors described in Table 1. A Keithley 2636A system sourcemeter and a Keithley 3706 system switch/MultiMeter were used to measure the  $I_{ds}$ - $V_{gs}$  (back sweep from +40 to -40V; in steps of 200 mV and constant  $V_{ds}$  of 2V) as a function of time. For each sample, the baseline responses of the sensors were recorded, first for 10 min in vacuum (30 mTorr), followed by 10 min exposure of the breath sample, and 10 min during the recovery of the sensors. To detect malfunctions and/or slight drifts of their baseline conditions, the sensors were calibrated daily by exposure to known concentrations of three calibration compounds, including 23.8 ppm isopropyl alcohol, 6.3 ppm trimethylbenzene, and 1.2 ppm 2-ethylhexanol before recording their resistance changes. Sensing features were read out and extracted from the time-dependent response of each sensor that related to voltage threshold ( $V_{th}$ ), hole mobility ( $\mu_h$ ), and source-drain current ( $I_{sd}$ ) at various back gate voltage ( $V_{gs}$ ) features. Breath patterns were obtained from the collective response of the sensors by applying discriminant factor analysis (DFA) and artificial neural networks (ANNs).

### ***Artificial Neural Networks (ANNs)***

ANNs is a machine-learning method inspired by neural networks in the human nervous system. It is based on a set of functions connecting the input (sensor features in this case) with the output (classification of samples to a specific disease). This relationship is achieved by optimizing certain parameters - connection weights, and amount of neurons (calculation centers) - to achieve the best classification from the inputs available. To reach the final mathematical binary classifying models to distinguish between the 4 groups of participants (LC, GC, Asthma/COPD, and negative to all), a 2-phase calculation process was used. The same steps were carried out for every available database (one per sensor (6) and possible binary classifier (6); a total of 36 models). Initially, a feature selection (FS) procedure was used to determine the sensing features with the greatest discriminative power to identify correctly the samples within a particular dataset. FS calculation was based on the Relief-F algorithm, which is a filter FS method. It operates by locating the nearest neighbors of a sample in its same class and other classes, giving priority to those features that distinguish better the nearest neighbors from different classes.<sup>58</sup> FS leads to the location of the best suited variables (sensing features in this case) to fulfil a second modeling phase, which leads to a reduced computational load and improved performance of further models.<sup>59</sup> The second part of the calculation is the modeling phase; using the features given by the FS test as independent variables, a

set of 36 multilayer perceptrons (MLPs), which are the most employed type of neural network, were designed and optimized. Based on non-linear interpolation for the calculations, MLPs are supervised algorithms, or in other words, tools that require target data (labelled data) to be trained properly.<sup>60</sup> In this case, the samples for every binary classifier were either tagged with a zero or a one depending on the group they belonged to. These labels become the dependent variable of each MLP classifier. MLPs, as their name suggests, are layered algorithms, possessing 3 kinds of layers (input, hidden and output). The input layer is formed by nodes that are only in charge of introducing the independent variables of the model (the sensing features selected during the FS phase). On the other hand, the hidden and output layers are formed by neurons, which are in charge of the calculations of the system. The number of neurons in the output layer are equivalent to the amount of dependent variables in the system (one in this case, as all the models are binary classifiers), whereas the hidden neuron number (HNN) (*vide infra*) and hidden layer amount must be correctly optimized (due to the size of the databases, the amount of hidden layers was set to one in all cases to avoid over-fitting).<sup>48</sup> Once the number of units (nodes and neurons) in each layer or network topology has been defined, the training process of the MLP can begin. The goal of this process is to optimize the weights that are contained within the model, there being as many weights as connections between units in neighboring layers (node-hidden neurons and hidden neurons-output neurons). The correct optimization of the weights is vital for a MLP to operate correctly, and this process is carried out during a series of training cycles or epochs. These cycles modify the values of the weights, with the goal set to lower the error of the estimation for a training dataset (in our case, to increase the correct classification ratio). This error can potentially reach zero if no precautions are followed, leading to an over-fit model. For this reason, the total database is initially divided into training and verification datasets, where the first is used to modify the weights, and the latter to verify that the MLP can perform well with samples that are not involved in this process. When the error for the verification dataset rises during 6 straight epochs, the training process is stopped, and the model is optimized as well as being not over-fit towards the training samples.<sup>61</sup> Besides the weights, there are other parameters that have to be correctly selected or optimized to reach a fully operational MLP. They are the HNN, the training and transfer functions, and a set of network parameters (Marquardt adjustment parameter (Lc), decrease factor for Lc (Lcd), and increase factor for Lc (Lci)), as follows. First, optimizing the HNN is crucial, as an excessive amount will most likely create over-fit models, and a low one will not allow the MLP to reach its maximum calculating potential; it has been optimized heuristically during



this research.<sup>62</sup> Second, the training function is in charge of the calculations that modify the weights. The one chosen here was the Levenberg-Marquardt (trainLM) function as it is the quickest algorithm for moderate-sized MLPs and has a memory reduction feature where the training dataset is large.<sup>62</sup> Third, the transfer function is in charge of introducing non-linearity in the calculations, as well as limiting the range of the values of the responses given by each neuron. The selected option was the sigmoid function, which limits the values between zero and one.<sup>54</sup> Finally, the Lc parameter is analogous to the learning coefficient in classic back-propagation algorithm;<sup>63</sup> its value decreases and increases with Lcd and Lci, respectively, until the modifications of Lc create a worse statistical performance.

### ***Discriminant Factor Analysis (DFA)***

DFA is a linear, supervised pattern recognition method that effectively reduces the multidimensional experimental data, in which the classes to be discriminated are defined before the analysis is performed. First, DFA was used for the selection of the sensors with the most relevant organic functionality out of a repertoire of 20, which was done by filtering out non-contributing sensors. DFA determines the linear combinations of the input variables (features being extracted from each sensor), so the variance in each class is minimized and that between classes maximized. The DFA output variables (*i.e.* canonical variables - CV) are achieved in mutually orthogonal dimensions; the first CV is the most powerful discriminating dimension. Leave-one-out cross-validation was conducted in calculating the classification success in terms of sensitivity, specificity and accuracy calculated as follows:

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

$$Accuracy = \frac{TP+TN}{Total} \quad (3)$$

Where TP stands for true positive, namely all positive samples correctly classified as positive; TN (true negative) - negative samples correctly classified as such; FP (false positive) - the negative sample incorrectly classified as positive; FN (false negative) - positive samples incorrectly classified as negative; and "total" is the total number of samples. Given k, the model was computed using k-1 training vectors. All possibilities of leave-one-sample-out were considered, and the classification

accuracy was estimated as the averaged performance over the k tests. Pattern recognition and data classification used MATLAB® (The MathWorks).

## **ASSOCIATED CONTENTS:**

### **Supporting Information**

Analytical values of the VOCs in the exhaled breath of patients and limit of detection of the examined sensors, statistical performance of every MLP model, summary of sensitivity / specificity / accuracy values achieved from DFA analysis of the samples, ROC curves for staging of LC and GC, are all available free of charge on the Internet: <http://pubs.acs.org>.

### **Corresponding Author**

\* H. Haick. Email: [hhossam@technion.ac.il](mailto:hhossam@technion.ac.il)

### **Acknowledgments**

This research received funding from the FP7-Health Program under the LCAOS (grant agreement no. 258868) and from the Horizon 2020 ICT Program under the SNIFFPHONE (grant agreement no. 644031). Clinical sample collection in Latvia was funded in part from the grant No. 305/2012 from the Latvian Council of Science. Fabrication was performed at the Micro-Nano Fabrication Unit (MNFU) of the Technion – Israel Institute of Technology. The authors acknowledge Mr. Alaa Gharra for assistance with statistical analysis and Dr. Viki Klopper for help in part of the schematics.

### **Competing financial interests**

The authors declare no competing interests.

### **References:**

1. Siegel, R. L.; Miller, K. D.; Jemal, A. Cancer Statistics, 2015. *CA: A Cancer J. Clin.* **2015**, *65*, 5-29.

2. Gibson, P. G.; Simpson, J. L. The Overlap Syndrome of Asthma and COPD: What Are Its Features and How Important Is It? *Thorax* **2009**, *64*, 728-735.
3. Broza, Y. Y.; Mochalski, P.; Ruzsanyi, V.; Amann, A.; Haick, H. Hybrid Volatolomics and Disease Detection. *Angew. Chem. Int. Ed.* **2015**, *54*, 11036-11048.
4. Haick, H.; Broza, Y. Y.; Mochalski, P.; Ruzsanyi, V.; Amann, A. Assessment, Origin, and Implementation of Breath Volatile Cancer Markers. *Chem. Soc. Rev.* **2014**, *43*, 1423-1449.
5. Nakhleh, M.; Broza, Y. Y.; Haick, H. Monolayer-Capped Gold Nanoparticles for Disease Detection from Breath. *Nanomedicine (London, UK)* **2014**, *9*, 1991-2002.
6. Amal, H.; Leja, M.; Funka, K.; Skapars, R.; Sivins, A.; Ancans, G.; Liepniece-Karele, I.; Kikuste, I.; Lasina, I.; Haick, H. Detection of Precancerous Gastric Lesions and Gastric Cancer Through Exhaled Breath. *Gut* **2016**, *65*, 400-407
7. Amal, H.; Leja, M.; Funka, K.; Lasina, I.; Skapars, R.; Sivins, A.; Ancans, G.; Kikuste, I.; Vanags, A.; Tolmanis, I.; Kirsners, A.; Kupcinskis, L.; Haick, H. Breath Testing as Potential Colorectal Cancer Screening Tool. *Int. J. Cancer* **2016**, *138*, 229-236.
8. Amal, H.; Shi, D.-Y.; Ionescu, R.; Zhang, W.; Hua, Q.-L.; Pan, Y.-Y.; Tao, L.; Liu, H.; Haick, H. Assessment of Ovarian Cancer Conditions from Exhaled Breath. *Int. J. Cancer* **2015**, *136*, E614-E622.
9. Ionescu, R.; Broza, Y.; Shaltieli, H.; Sadeh, D.; Zilberman, Y.; Feng, X.; Glass-Marmor, L.; Lejbkowitz, I.; Müllen, K.; Miller, A.; Haick, H. Detection of Multiple Sclerosis from Exhaled Breath Using Bilayers of Polycyclic Aromatic Hydrocarbons and Single-Wall Carbon Nanotubes. *ACS Chem. Neurosci.* **2011**, *2*, 687-693.
10. Tisch, U.; Schlesinger, I.; Ionescu, R.; Nassar, M.; Axelrod, N.; Robertman, D.; Tessler, Y.; Marmur, A.; Aharon-Peretz, J.; Haick, H. Detection of Alzheimer's and Parkinson's Diseases from Exhaled Breath Using Nanomaterial-Based Sensors. *Nanomedicine (London, UK)* **2013**, *8*, 43-56.
11. Konvalina, G.; Haick, H. The effect of Humidity on Nanoparticle-Based Chemiresistors: A Comparison Between Synthetic and Real-World Samples. *ACS Appl. Mater. Interfaces* **2012**, *4*, 317-325.
12. Nakhleh, M. K.; Amal, H.; Awad, H.; Gharra, A.; Abu-Saleh, N.; Jeries, R.; Haick, H.; Abassi, Z. Sensor Arrays Based on Nanoparticles for Early Detection of Kidney Injury by Breath Samples. *Nanomedicine (New York, NY, US)* **2014**, *19*, 00320-00327.
13. Nakhleh, M. K.; Jeries, R.; Gharra, A.; Binder, A.; Broza, Y. Y.; Pascoe, M.; Dheda, K.; Haick, H. Detecting Active Pulmonary Tuberculosis with a Breath Test using Nanomaterial-Based Sensors. *Eur. Respir. J.* **2014**, *43*, 1522-1525.
14. Xu, Z. q.; Broza, Y. Y.; Ionsecu, R.; Tisch, U.; Ding, L.; Liu, H.; Song, Q.; Pan, Y. Y.; Xiong, F. X.; Gu, K. S.; Sun, G. P.; Chen, Z. D.; Leja, M.; Haick, H. A Nanomaterial-Based Breath Test for Distinguishing Gastric Cancer from Benign Gastric Conditions. *Br. J. Cancer* **2013**, *108*, 941-950.

15. Hakim, M.; Billan, S.; Tisch, U.; Peng, G.; Dvorkind, I.; Marom, O.; Abdah-Bortnyak, R.; Kuten, A.; Haick, H. Diagnosis of Head&Neck Cancer from Exhaled Breath. *Br. J. Cancer* **2011**, *104*, 1649-1655.
16. Amann, A.; Mochalski, P.; Ruzsanyi, V.; Broza, Y. Y.; Haick, H. Assessment of the Exhalation Kinetics of Volatile Cancer Biomarkers Based on their Physicochemical Properties. *J. Breath. Res.* **2014**, *8*, 016003.
17. Hakim, M.; Broza, Y. Y.; Barash, O.; Peled, N.; Phillips, M.; Amann, A.; Haick, H. Volatile Organic Compounds of Lung Cancer and Possible Biochemical Pathways. *Chem. Rev.* **2012**, *112*, 5949-5966.
18. Konvalina, G.; Haick, H. Sensors for Breath Testing: From Nanomaterials to Comprehensive Disease Detection. *Acc. Chem. Res.* **2013**, *47*, 66-76.
19. Phillips, M.; Cataneo, R. N.; Cummin, A. R. C.; Gagliardi, A. J.; Gleeson, K.; Greenberg, J.; Maxfield, R. A.; Rom, W. N. Detection of Lung Cancer with Volatile Markers in the Breath. *Chest* **2003**, *123*, 2115-2123.
20. Simenhoff, M. L.; Burke, J. F.; Saukkonen, J. J.; Ordinario, A. T.; Doty, R.; Dunn, S. Biochemical Profile of Uremic Breath. *New. Eng. J. Med.* **1977**, *297*, 132-135.
21. Phillips, M. Detection of Carbon Disulfide in Breath and Air: A Possible New Risk Factor for Coronary Artery Disease. *Int. Archiv. Occup. Env. Health* **1992**, *64*, 119-123.
22. Wang, C.; Sahay, P. Breath Analysis Using Laser Spectroscopic Techniques: Breath Biomarkers, Spectral Fingerprints, and Detection Limits. *Sensors* **2009**, *9*, 8230-8262.
23. Risby, T. H.; Solga, S. F. Current Status of Clinical Breath Analysis. *Appl. Phys. B* **2006**, *85*, 421-426.
24. Miekisch, W.; Schubert, J. K. From Highly Sophisticated Analytical Techniques to Life-Saving Diagnostics: Technical Developments in Breath Analysis. *TrAC Trends Anal. Chem.* **2006**, *25*, 665-673.
25. Martin, A. N.; Farquar, G. R.; Jones, A. D.; Frank, M. Human Breath Analysis: Methods for Sample Collection and Reduction of Localized Background Effects. *Anal. Bioanal. Chem.* **2010**, *396*, 739-750.
26. Bajtarevic, A.; Ager, C.; Pienz, M.; Klieber, M.; Schwarz, K.; Ligor, M.; Ligor, T.; Filipiak, W.; Denz, H.; Fiegl, M. Noninvasive Detection of Lung Cancer by Analysis of Exhaled Breath. *BMC Cancer* **2009**, *9*, 348.
27. Machado, R. F.; Laskowski, D.; Deffenderfer, O.; Burch, T.; Zheng, S.; Mazzone, P. J.; Mekhail, T.; Jennings, C.; Stoller, J. K.; Pyle, J. Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath. *Am. J. Resp. Crit. Care Med.* **2005**, *171*, 1286-1291.
28. Fleischer, M.; Simon, E.; Rumpel, E.; Ulmer, H.; Harbeck, M.; Wandel, M.; Fietzek, C.; Weimar, U.; Meixner, H. Detection of Volatile Compounds Correlated to Human Diseases Through Breath Analysis with Chemical Sensors. *Sens. Actuat. B* **2002**, *83*, 245-249.

29. Peled, N.; Hakim, M.; Tisch, U.; Bunn, P. A. J. R.; Miller, Y. E.; Kennedy, T. C.; Mattei, J.; Mitchell, J. D.; Weyant, M. J.; Hirsch, F. R.; Haick, H. Non-Invasive Breath Analysis of Pulmonary Nodules. *J. Thorac. Oncol.* **2012**, *7*, 1528-1533.
30. Vishinkin, R.; Haick, H. Nanoscale Sensor Technologies for Disease Detection via Volatolomics. *Small* **2015**, *11*, 6142-6164.
31. Broza, Y. Y.; Haick, H. Nanomaterial-Based Sensors for Detection of Disease by Volatile Organic Compounds. *Nanomedicine (London, UK)* **2013**, *8*, 785-806.
32. Park, H.; Dan, Y.; Seo, K.; Yu, Y. J.; Duane, P. K.; Wober, M.; Crozier, K. B. Filter-Free Image Sensor Pixels Comprising Silicon Nanowires with Selective Color Absorption. *Nano Lett.* **2014**, *14*, 1804-1809.
33. Wang, B.; Cancilla, J. C.; Torrecilla, J. S.; Haick, H. Artificial Sensing Intelligence with Silicon Nanowires for Ultraselective Detection in the Gas Phase. *Nano Lett.* **2014**, *14*, 933-938.
34. Paska, Y.; Haick, H. Interactive Effect of Hysteresis and Surface Chemistry on Gated Silicon Nanowire Gas Sensors. *ACS Appl. Mater. Interfaces* **2012**, *4*, 2604-2617.
35. Paska, Y.; Stelzner, T.; Tisch, U.; Assad, O.; Christiansen, S.; Haick, H. Molecular Gating of Silicon Nanowire Field Effect Transistors with Nonpolar Analytes. *ACS Nano* **2012**, *6*, 335-345.
36. Paska, Y.; Stelzner, T.; Christiansen, S.; Haick, H. Enhanced Sensing of Nonpolar Volatile Organic Compounds by Silicon Nanowire Field Effect Transistors. *ACS Nano* **2011**, *5*, 5620-5626.
37. Paska, Y.; Haick, H. Controlling Properties of Field Effect Transistors by Intermolecular Cross-Linking of Molecular Dipoles. *Appl. Phys. Lett.* **2009**, *95*, 233103/1-3.
38. Lou, L.; Zhang, S.; Park, W.-T.; Tsai, J. M.; Kwong, D.-L.; Lee, C. Optimization of NEMS Pressure Sensors with a Multilayered Diaphragm Using Silicon Nanowires as Piezoresistive Sensing Elements. *J. Micromechan. Microeng.* **2012**, *22*, 055012.
39. Demami, F.; Ni, L.; Rogel, R.; Salaun, A.-C.; Pichon, L. Silicon Nanowires based Resistors as Gas Sensors. *Sens. Actuat. B* **2012**, *170*, 158-162.
40. Yan, Q.; Wang, Z.; Zhang, J.; Peng, H.; Chen, X.; Hou, H.; Liu, C. Nickel Hydroxide Modified Silicon Nanowires Electrode for Hydrogen Peroxide Sensor Applications. *Electrochim. Acta* **2012**, *61*, 148-153.
41. Halpern, J. M.; Wang, B.; Haick, H. Controlling the Sensing Properties of Silicon Nanowires via the Bonds Nearest to the Silicon Nanowire Surface. *ACS Appl. Mater. Interfaces* **2015**, *7*, 11315-11321.
42. Ahmed, L. B.; Naama, S.; Keffous, A.; Hassen-Bey, A.; Hadjersi, T. H<sub>2</sub> sensing properties of modified silicon nanowires. *Progress in Natural Science: Materials International* **2015**.
43. Kim, Y.; Kim, H. J.; Kim, J. H.; Choi, D. G.; Choi, J. H.; Jung, J. Y.; Jeon, S.; Lee, E. S.; Jeong, J.-H.; Lee, J. Rapid Low-Temperature 3D Integration of Silicon Nanowires on Flexible Substrates. *Small* **2015**, *11*, 3995-4001.

44. Wang, H.; Fan, P. H.; Tong, B.; Dong, Y. P.; Ou, X. M.; Li, F.; Zhang, X. H. Hydrogen-Terminated Si Nanowires as Label-Free Colorimetric Sensors in the Ultrasensitive and Highly Selective Detection of Fluoride Anions in Pure Water Phase. *Adv. Funct. Mater.* **2015**, *25*, 1506–1510.
45. Wang, B.; Haick, H. Effect of Functional Groups on the Sensing Properties of Silicon Nanowires Toward Volatile Compounds. *ACS Appl. Mater. Interfaces* **2013**, *5*, 2289-2299.
46. Wang, B.; Haick, H. Effect of Chain Length on the Sensing of Volatile Organic Compounds by Means of Silicon Nanowires. *ACS Appl. Mater. Interfaces* **2013**, *5*, 5748-5756.
47. Ermanok, R.; Assad, O.; Zigelboim, K.; Wang, B.; Haick, H. Discriminative Power of Chemically Sensitive Silicon Nanowire Field Effect Transistors to Volatile Organic Compounds. *ACS Appl. Mater. Interfaces* **2013**, *5*, 11172-11183.
48. Bashouti, M. Y.; Stelzner, T.; Berger, A.; Christiansen, S.; Haick, H. Covalent Attachment of Alkyl Functionality to 50 nm Silicon Nanowires Through Chlorination / Alkylation Process. *J. Phys. Chem. C* **2009**, *113*, 14823-14828.
49. Bashouti, M. Y.; Stelzner, T.; Berger, A.; Christiansen, S.; Haick, H. Chemical Passivation of Silicon Nanowires with C1–C6 Alkyl Chains Through Covalent Si–C Bonds. *J. Phys. Chem. C* **2008**, *112*, 19168-19172.
50. Haick, H.; Ambrico, M.; Ligonzo, T.; Cahen, D. Discontinuous Molecular Films can Control Metal/Semiconductor Junctions. *Adv. Mater.* **2004**, *16*, 2145-2151.
51. Wang, B.; Huynh, T.-P.; Wu, W.; Hayek, N.; Do, T. T.; Cancilla, J. C.; Torrecilla, J. S.; Nahid, M. M.; Colwell, J. M.; Gazit, O. M.; Puniredd, S. R.; McNeill, C. R.; Sonar, P.; Haick, H. Highly Sensitive Ambipolar Field Effect Transistor-Based Diketopyrrolopyrrole Copolymer for Selective Detection and Discrimination of Xylene Isomers. *Adv. Mater.* **2016**, *28*, 4012-4018.
52. Team, T. N. L. S. T. R. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *N. Engl. J. Med.* **2013**, *368*, 1980-1991.
53. Pasechnikov, V.; Chukov, S.; Fedorov, E.; Kikuste, I.; Leja, M. Gastric Cancer: Prevention, Screening and Early Diagnosis. *World J. Gastroenterol.* **2014**, *20*, 13842-13862.
54. Knoerzer, K.; Juliano, P.; Roupas, P.; Versteeg, C.: *Innovative Food Processing Technologies: Advances in Multiphysics Simulation*; Wiley-Blackwell: Oxford (UK), 2011.
55. Dunteman, G. H.: *Introduction to multivariate analysis*; Sage Publications: Thousand Oaks, CA, 1984.
56. Assad, O.; Leshansky, A. M.; Wang, B.; Stelzner, T.; Christiansen, S.; Haick, H. Spray-Coating Route for Highly Aligned and Large-Scale Arrays of Nanowires. *ACS Nano* **2012**, *6*, 4702-4012.

57. Amal, H.; Leja, M.; Broza, Y. Y.; Tisch, U.; Funka, K.; Liepniece-Karele, I.; Skapars, R.; Xu, Z. Q.; Liu, H.; Haick, H. Geographical Variation in the Exhaled Volatile Organic Compounds. *J. Breath Res* **2013**, *7*, 047102.
58. Wu, B.; Chen, C.; Kechadi, T. M.; Sun, L. A Comparative Evaluation of Filter-Based Feature Selection Methods for Hyper-Spectral Band Selection. *Int. J. Rem. Sens.* **2013**, *34*, 7974-7990.
59. Chandrashekar, G.; Sahin, F. A Survey on Feature Selection Methods. *Comput. Elect. Eng.* **2014**, *40*, 16-28.
60. Basheer, I. A.; Hajmeer, M. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *J. Microbiol. Meth.* **2000**, *43*, 3-31.
61. Torrecilla, J. S.; Tortuero, C.; Cancilla, J. C.; Díaz-Rodríguez, P. Estimation with Neural Networks of the Water Content in Imidazolium-Based Ionic Liquids Using Their Experimental Density and Viscosity Values. *Talanta* **2013**, *113*, 93-98.
62. Cancilla, J. C.; Díaz-Rodríguez, P.; Matute, G.; Torrecilla, J. S. The Accurate Estimation of Physicochemical Properties of Ternary Mixtures Containing Ionic Liquids *via* Artificial Neural Networks. *Phys. Chem. Chem. Phys.* **2015**, *17*, 4533-4537.
63. Palancar, M. C.; Aragon, J. M.; Torrecilla, J. S. pH-Control System Based on Artificial Neural Networks. *Ind. Eng. Chem. Res.* **1998**, *37*, 2729-2740.

## ToC Graphics:

