



## Depletion of wild-type target enhances the hybridization-based sensitivity of low-abundant mutation detection by reference capture probes

Rebekka Van Hoof<sup>a,b,c,1</sup>, Michal Szymonik<sup>a,2,3</sup>, Stefanos K. Nomidis<sup>a,c,4</sup>, Karen Hollanders<sup>a,5</sup>, An Jacobs<sup>a,6</sup>, Inge Nelissen<sup>a,7</sup>, Patrick Wagner<sup>c,8</sup>, Jef Hooyberghs<sup>a,b,d,\*,9</sup>

<sup>a</sup> Flemish Institute for Technological Research (VITO), Health Unit, Boeretang 200, 2400 Mol, Belgium

<sup>b</sup> Data Science Institute, Theoretical Physics, Hasselt University, Campus Diepenbeek, 3590 Diepenbeek, Belgium

<sup>c</sup> Laboratory for Soft Matter and Biophysics, KU Leuven, Celestijnenlaan 200 D, 3001 Leuven, Belgium

<sup>d</sup> Flemish Institute for Technological Research (VITO), Data Science, Boeretang 200, 2400 Mol, Belgium

### ARTICLE INFO

#### Keywords:

Hybridization-based nucleic acid mutation sensors  
DNA thermodynamics  
Target depletion  
Single nucleotide variant  
Microarray technology  
Reference capture probes

### ABSTRACT

Nucleic acids duplex formation via hybridization is a crucial reaction in many processes and application across different disciplines. In life sciences the detection of mutations is an important application for which hybridization is used, e.g. in diagnostics via single-nucleotide variants (SNVs). This paper deals with the physico-chemical aspects of hybridization-based detection of low-abundance mutations, which is challenging due to unavoidable competitive hybridization of high-abundant wild type sequence with the low-abundant variants. We apply two experimental methods based on theoretical hybridization models to show how sensing of DNA mutation can be significantly improved. This is implemented on two SNV biomarkers for which we first select a reference capture probe. This is a probe designed to match neither the wild type nor the SNV sequence, but to have an equal affinity to the wild-type as the SNV-matching probe. This allows the mutation-specific signal to be expressed as a ratiometric quantity, leading to increased assay robustness. Secondly, we selectively deplete the wild-type species by introducing an excess of wild-type-specific capture probes, and account for these depletion effects in the theoretical model. We demonstrate the detection of 0.05% mutant species in a wild-type background, which is an improvement of an order of magnitude in the limit of detection in comparison with the no-depletion case. This sensitivity is comparable with digital PCR results, showing performance suitable for e.g. clinical applications in liquid biopsy context. The principles of this work apply to a wide range of hybridization-based DNA biosensing technologies, irrespective of the underlying transducer principle.

**Abbreviations:** COSMIC, Catalog, ue Of Somatic Mutations In Cancer; ctDNA, circulating tumor DNA; dPCR, digital PCR; EGFR, epidermal growth factor receptor; I, intensity; KRAS, Kirsten rat sarcoma viral oncogene homolog; LoD, limit of detection; LRG, Locus Reference Genomic; Mut, mutant; NGS, next generation sequencing; P, probe; PCR, polymerase chain reaction; Ref, reference; S, signal; S<sub>0</sub>, baseline signal; SNV, single-nucleotide variant; T, target; WT, wild-type.

\* Corresponding author at: Flemish Institute for Technological Research (VITO), Health Unit, Boeretang 200, 2400 Mol, Belgium.

E-mail address: [jef.hooyberghs@vito.be](mailto:jef.hooyberghs@vito.be) (J. Hooyberghs).

<sup>1</sup> 0000-0002-9982-2826

<sup>2</sup> These authors contributed equally.

<sup>3</sup> 0000-0002-7107-0976

<sup>4</sup> 0000-0002-4635-8433

<sup>5</sup> 0000-0002-8546-0174

<sup>6</sup> 0000-0002-7178-0185

<sup>7</sup> 0000-0002-1248-3463

<sup>8</sup> 0000-0002-4028-3629

<sup>9</sup> 0000-0003-3781-9645

<https://doi.org/10.1016/j.snb.2022.132175>

Received 24 February 2022; Received in revised form 3 June 2022; Accepted 5 June 2022

Available online 7 June 2022

0925-4005/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

### 1.1. Biosensing of low-abundant mutations and its relevance for applications

Hybridization of single-stranded nucleic acids to form a duplex is a reaction which is at the heart of many processes and applications in physics, chemistry, biology and medicine. In the latter two disciplines the detection of mutations is an important topic because of their functional relevance in life, and hybridization is one of the techniques applied for this task next to variants of polymerase chain reaction and sequencing [1–8]. Each of the techniques has its own advantages: for hybridization, these are e.g. the possibility to be used in a multitude of miniaturized biosensor devices and the ability to operate in a wide variety of conditions [9–13]. In our work we use microarrays for the technological implementation because of their maturity and highly parallelized fluorescent read-out system. The results are general since the principles apply to all hybridization-based devices, independent of the specific transducer principle such as fluorescent, thermal, impedimetric, surface plasmon resonance, mass-based etc. This paper deals with hybridization-based biosensing of mutations which are present in low abundance. This is at the same time highly relevant e.g. in diagnostics via single-nucleotide variants (SNVs) and challenging from technical-scientific point of view. The research focus of this paper is on a significant experimental progress in the chemical sensing of DNA mutations, not on clinical diagnostics. However, we will start by briefly addressing both and end this subsection with a clear problem statement and specific goals.

Hybridization-based low-abundance sensing is inevitably faced with competitive hybridization of various target sequences, i.e. the high-abundant wild type sequence and the low-abundant nucleotide variants [5,14]. A commonly applied method for the enhancement of SNVs detection is the enrichment of their respective share in the target sample, either by selective amplification of SNVs or by depletion of wild-type sequences [15,16]. Different strategies for selective amplification of the mutated sequence have been explored based on differences in melting temperature between a “blocker sequence” and the SNV in comparison to the wild-type sequence [17]. Examples include 3'-modified primers, locked-nucleic-acid or peptide-nucleic-acid clamps [18–20]; thermodynamically designed simulation-guided DNA probes with a very high affinity for SNVs [21] and selective exonuclease with different digestive affinity towards mismatched sequences [22]. In our work, we implement a different strategy where wild-type depletion is induced during the hybridization-based read-out. For this, we employ an excess of probe sequences to capture wild-type target sequences, in combination with reference probes for the selective detection of low abundant SNV sequences to enhance the hybridization-based detection limit.

In this approach we use the thermodynamic analysis of the relative free energy penalties of various single-nucleotide sequence mismatches [23–25] to design a dedicated set of capture probe sequences with favorable thermodynamic properties both towards detection sensitivity and dynamic range of the measurement device. This approach has previously been applied to the analysis of human HIV samples [7], and strain typing of tuberculosis [26]. Moreover, the detection of 1% Kirsten rat sarcoma viral oncogene homolog (*KRAS*) oncogene SNVs in a wild-type background has been shown in formalin-fixed, paraffin-embedded human tissue samples [2]. We recently extended this theoretical model to account for target depletion effects. Originally intended to account for possible non-idealities in the sensing readout, the analysis indicated that depletion may actually lead to significant improvements in assay sensitivity under certain conditions [27]. The extended theoretical model predicted that selective depletion of the wild-type target may lead to a significant improvement in detection sensitivity for low-abundant SNVs. In the current work, we present the results of an experimental realization of this depletion effect. The goal is

to demonstrate the improvement of the sensing capability and to illustrate the validity, we chose a clinically relevant example in the implementation: SNVs for lung cancer in the rapidly emerging application of analysis in minimally-invasive liquid biopsy samples [28]. In liquid biopsies, circulating tumor DNA (ctDNA) is profiled in blood, minimizing patient discomfort, risk and costs, and allowing for more frequent and less heterogeneous sampling. It is a technology of growing importance which, in early stages of tumor development, requires the detection of less than 0.1% of mutated tumor DNA in a background of wild-type DNA [29,30].

We set out to test the effect of wild-type target depletion on the detection performance of SNVs relevant to non-small cell lung cancer, which represents 85% of all diagnosed lung cancer cases [31]. Mutations in the epidermal growth factor receptor (*EGFR*) and *KRAS* genes are important biomarkers to determine the course of therapy [32]. We selected one SNV on each gene as example biomarkers relevant for the selection and follow-up of treatment in lung cancer, these targets are shown in Table 1, where the coordinates of the DNA mutation and the amino acid substitution of the resulting protein are indicated.

### 1.2. Problem statement and objectives

The microarrays used in this work are composed of spots, each functionalized with an equal amount of a specific ssDNA sequence, termed a probe, as shown in Fig. 1. In the sample being studied, fluorescently labeled ssDNA is present either as a pure wild-type target or as a mixture of wild-type and mutant targets (*Twt* and *Tmut*), with *Tmut* typically at a low concentration relative to *Twt*, see Fig. 2. During hybridization, a fraction of the target binds to the probes on the array, after which the total amount of target (i.e. the sum of *Twt* and *Tmut*) hybridized to each individual spot is measured optically to produce a unitless intensity (*I*) value. The probability of target-probe hybridization is highest for a fully Watson-Crick complementary target-probe pair, and it is lower but often non-zero for partially-mismatched probes. For instance, *Twt* has the highest affinity towards its fully complementary probe (*Pwt*) and hence contributes most to the intensity of a *Pwt* array location. Quantitatively, each target-probe affinity is determined by the Gibbs free energy  $\Delta G$  of the corresponding target-probe duplex: the  $\Delta G$  value corresponds to the affinity and binding probability.

Due to the high fractional concentration of *Twt* in the sample, the relative contribution from *Twt* binding to *Pmut* can be large, despite the sequence mismatch between them. This produces a large background signal and means variations in the concentration of *Twt* would interfere with determining the presence of *Tmut* from *Pmut* intensity alone.

The first aim of this study is to experimentally show that this effect can be counteracted and sensitive, robust SNV detection is enabled by selecting appropriate probe sets. To this end, we use the theoretically proposed design of Nomidis et al. [27], which is sketched in Fig. 2, where a reference probe *Pref* is used next to the classical probes *Pwt* and *Pmut*. A crucial point is that the sequence of *Pref* is chosen such that  $\Delta G_{Twt+Pref} = \Delta G_{Twt+Pmut}$ , i.e. the affinity of the reference and the mutant probes towards the wild-type target are equal under the experimental conditions. Hence, in the absence of mutant target, i.e. a pure wild-type

**Table 1**

Details of the target lung cancer biomarker SNVs and their identification number in the Catalogue Of Somatic Mutations In Cancer (COSMIC) and Locus Reference Genomic (LRG) databases [33,34].

Variant name	COSMIC mutation ID	Gene name (LRG record ID)	DNA mutation, coding DNA coordinates	Amino acid mutation, protein coordinates
<i>EGFR</i> T790M	COSM21943	<i>EGFR</i> (LRG_304)	c.034 G>A	p.T790M
<i>KRAS</i> G12S	COSM517	<i>KRAS</i> (LRG_344)	c.02369 C>T	p.G12S

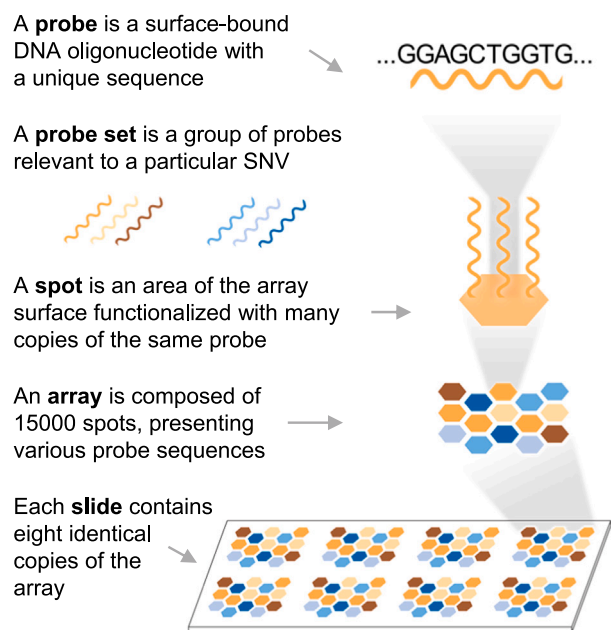


Fig. 1. Overview of the terms used throughout this work relating to microarray experiments. SNV: single-nucleotide variant, A: adenine, C: cytosine, G: guanine, T: thymine.

sample, array spots containing *Pref* and *Pmut* will produce equal read-out intensity:  $I_{Pmut} = I_{Pref}$ . When mutant target is present in the sample it will have largest affinity towards *Pmut* resulting in  $I_{Pmut} > I_{Pref}$ . Consequently, the ratio  $I_{Pmut} / I_{Pref}$  can be used as an indicator for the presence of *Tmut*. A more complete treatment of the hybridization thermodynamics is given in the results section.

The second, and principal aim of the present study is to experimentally verify the theoretical prediction of Nomidis et al. [27] that depletion of the wild-type target will yield improvements in the limit of

detection of SNVs.

## 2. Materials and methods

### 2.1. Target DNA preparation

Sequence-verified, double-stranded DNA fragments were purchased from IDT (gBlocks Gene Fragments, Integrated DNA Technologies, Leuven, Belgium). The target templates for PCR were 500 bp fragments of the human *EGFR* gene containing the T790M mutation, or of the human *KRAS* gene containing the G12S mutation, as well as their wild-type equivalents (*EGFR* wt, *KRAS* wt). For each gene, mutant and wild-type gBlocks were mixed to produce a dilution series between 3.2% and 0.0032% mutant DNA in a wild-type background, with a constant total DNA concentration of 67 pM. For details of sample preparation see [Supporting Information S1.1](#).

### 2.2. PCR amplicon generation

Full details of the amplicon preparation and primer sequences used are given in [Supporting Information S1.2](#). In brief, diluted mixtures of the mutant and wild-type gene fragments were amplified by PCR and purified. The amplicons were designed to be around 130 base pairs to match the typical length of ctDNA fragments [35]. Each reverse primer contained a common "barcode" sequence not present on the template, as well as a 5' phosphate modification (see [Supporting Information, Table S3](#)). The 5'-phosphorylated strands of the purified amplicons were then selectively digested with lambda exonuclease to produce single-stranded DNA, and used without further purification. Exonuclease digestion was confirmed by gel electrophoresis see [Supporting Information S2](#).

### 2.3. Microarray probe selection and design

For the experimental selection of reference capture probes, we used microarrays containing all possible single and dual mismatches against

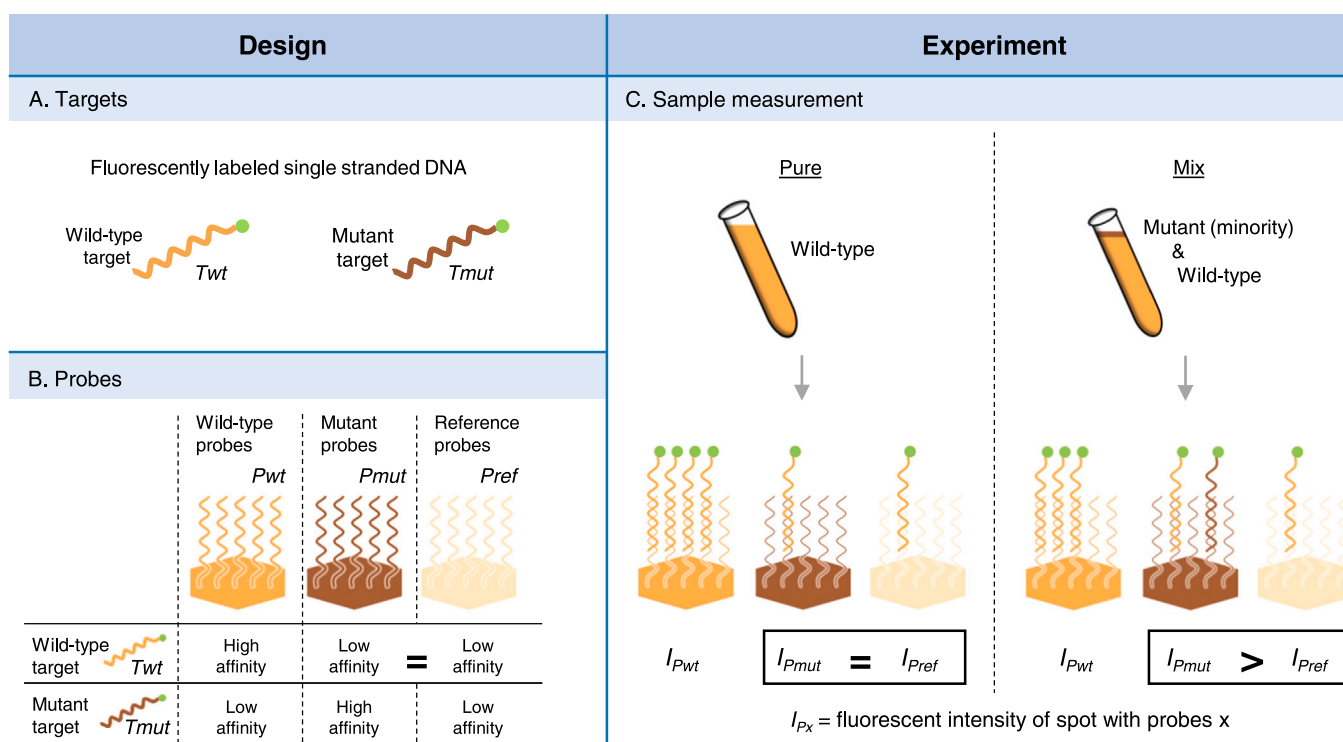


Fig. 2. Schematic representation of the design and the experimental readout of mutation detection with a reference probe.

the *KRAS* and *EGFR* wild-type sequences in the immediate vicinity of the mutations of interest. From this comprehensive set of probe options, the most suitable was chosen for each of the two genes (see Section 3.1 and Fig. 3).

Next, the limit of detection for each SNV was experimentally determined using a microarray designed with an over-abundance of *Pwt* spots. Since only three different probe sequences are used for the detection of each SNV, we can allocate in our microarray design the majority of the 15,000 spots to *Pwt*, while keeping 20 technical replicates of all other probes for statistical robustness. Further, the SNV probe sets for both *EGFR* and *KRAS* are included on two arrays, but with each designed to deplete only one of the two wild-type targets, leading to the scheme shown in Fig. 4.

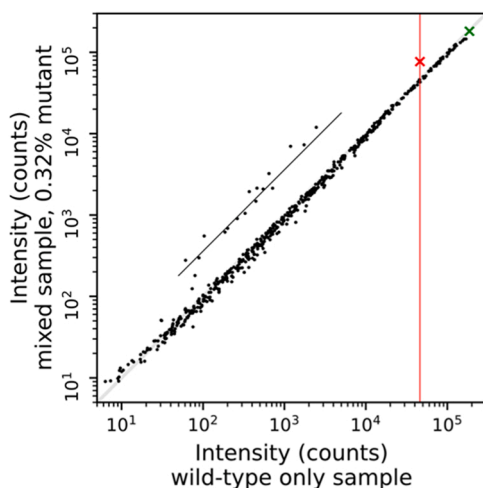
#### 2.4. DNA microarray experiment

Custom 8 × 15 K Agilent microarray slides (Agilent Technologies, Santa Clara, California, USA, #G2509F) were used. The array hybridization, washing and readout were carried out with the Gene Expression Microarrays Hybridization Kit (Agilent) according to manufacturer guidelines, and previously described by Willems et al. [2]. Digested *KRAS* and *EGFR* amplicons were present in the hybridization mixture at either 0.01 nM or 1 nM, depending on the condition tested. The hybridization mixture additionally contained 0.05 μM Cy-3 labeled barcode oligonucleotides, to avoid direct labeling of the target, as in previous work. Each slide contained 8 identical designs of the 15,000 spot array, and each array was exposed to a different sample mixture, allowing a range of eight different mutant concentrations to be measured on one slide simultaneously, see Supplementary Information S1.3.

The array images were processed using Agilent Feature Extraction Software (GE1 v5 95 Feb07). Spot fluorescence intensities were background corrected via subtraction of the global intensity minimum. The median value of replicate spot intensities was used in calculations of Signal values.




#### 2.5. Digital polymerase chain reaction

A QX200 Droplet Digital PCR System (Bio-Rad Laboratories,






**Fig. 3.** Distribution of probe fluorescent intensities in a sample of wild-type only target (x-axis) and for a mixture of mutant and wild-type target (y-axis) for the *EGFR* T790M mutant. The collection of probes contains single- and double-point mismatches with respect to the wild-type target sequence. The probes of the deviation branch have a higher relative affinity towards the *EGFR* T790M mutant. The green and red cross represent the wild-type specific probe and the mutant-specific probe, respectively. The intersection between the diagonal and the red vertical line indicates reference probe candidates.

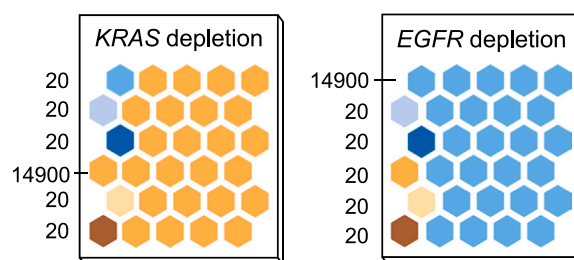
#### Probe set *KRAS* G12S mutation

	<i>Pwt</i>	GTTGGAGCTGGTGGCGTAGGCCAA
	<i>Pmut</i>	GTTGGAGCT <u>A</u> GTTGGCGTAGGCCAA
	<i>Pref</i>	GTTGG <u>I</u> GCTGGTGGCGT <u>G</u> GGCAA

#### Probe set *EGFR* T790M mutation

	<i>Pwt</i>	AAGGGCATGAGCTGCGTGATGAG
	<i>Pmut</i>	AAGGGCATGAGCTG <u>C</u> ATGATGAG
	<i>Pref</i>	AAGGGCATGAGC <u>G</u> GCGTG <u>G</u> TGAG

#### Number of replicate spots on array



**Fig. 4.** Microarray probe sets used in the experiment containing the wild-type (*Pwt*), mutant (*Pmut*) and reference (*Pref*) probe sequence for each mutant of interest. Mismatches against the wild-type target are highlighted in the *Pmut* and *Pref* sequences. The two microarray designs contain the same probe sets, but each array contains an excess of *Pwt* probes for one of the two SNVs of interest.

Hercules, California, USA, #1864001) was used with validated assays for the *EGFR* T790M and *KRAS* G12S mutations (Bio-Rad) according to the manufacturer's instructions to measure the same DNA mixtures as for the microarray experiment. Details of the digital PCR protocol are given in Supplementary Information S1.4.

### 3. Results and discussion

Using the design presented in Section 2.3, we determined the sensitivity of SNV detection in standard situations, i.e. in the absence of wild-type target depletion, and in situations where strong wild-type target depletion was intentionally induced. We quantitatively show the depletion-induced improvement that can be expected from theory, this is inherently limited by the physico-chemical reality of hybridization, and compared it with the experimental results. Finally, a clinically applied method for the analysis of liquid biopsy samples, digital PCR (dPCR), was used as a benchmarking technology, see Section 2.5 above.

#### 3.1. Reference capture probe selection and data analysis for SNV detection

We initially carried out a set of microarray experiments using probe sets containing all possible mismatches against the *EGFR* wt and *KRAS* wt gene fragments. In Fig. 3 the intensity of each spot on such an array when exposed to a sample containing only wild-type targets is plotted, versus a sample containing a mixture of a mutant and wild-type targets. The range of free energy penalties associated with specific sequence mismatches translates to a broad range of signal intensities for the

various probes. Intensities corresponding to probes that are “neutral” to the binding of the mutant species are located on the identity diagonal of the plot, while those with a stronger binding affinity for the mutant are located above this line. This “branch structure” forms the basis of the analysis used here for determining the presence of mutant species.

In Nomidis et al. [27] we theoretically described how, based on such a dataset, a reduced probe set can be designed which enables the detection of SNVs using only three probe sequences per mutation of interest. The probes that are perfectly complementary to the wild-type target (*Pwt*) and to the mutant (*Pmut*) are known. Here we experimentally selected the third “reference” probe (*Pref*), designed such that  $\Delta G_{Twt+Pref} \approx \Delta G_{Twt+Pmut}$ . The probe sequence yielding a fluorescence intensity closest to those of *Pmut* when hybridized with *Twt* was chosen as the *Pref* probe (one representative is shown in Fig. 3, red vertical line). All probe sequences used throughout this work are shown in Fig. 4.

Having selected appropriate reference probes, we define the mutant detection signal (*S*) as in previous work [27].

$$S \equiv \ln \frac{I_{Pmut}}{I_{Pref}} = \ln \frac{\theta_{Pmut}}{\theta_{Pref}} \quad (1)$$

where the measured intensity of each probe  $I_p$  is proportional to the fraction  $\theta_p$  of targets bound by a probe. For a sample containing only *Twt*, we expect by design  $\theta_{Pref} = \theta_{Pmut}$  leading to  $S_0 = 0$ ; a baseline signal of zero. In the case of a sample containing both *Twt* and *Tmut*, high-affinity binding of *Tmut* to *Pmut* leads to an increase in  $\theta_{Pmut}$  while  $\theta_{Pref}$  remains effectively constant, yielding an increase in the signal *S*. Eq. 2 describes how the detection signal depends on the relative concentration of the target species and on the free energy penalty associated with the mismatched residue in the sequence,  $\Delta\Delta G_{Pmut}$ . This is the difference ( $\Delta$ ) in hybridization free energy  $\Delta G$  between the mutant probe and its two targets: (*Pmut* + *Tmut*) versus (*Pmut* + *Twt*).

$$\text{Case1, no depletion : } S_1 = \ln \left\{ 1 + \frac{c_{Tmut}}{c_{Twt}} \exp\left(\frac{\Delta\Delta G_{Pmut}}{RT}\right) \right\} \quad (2)$$

where  $c_{Twt}$  and  $c_{Tmut}$  are the concentrations of the wild-type and mutant targets, *R* is the gas constant and *T* the absolute temperature. In the case of a large excess of *Pwt*, leading to wild-type target depletion, the same derivation yields Eq. (3), with the additional term  $\Delta\Delta G_{Pwt}$  being the difference in hybridization free energy  $\Delta G$  between the wild-type probe and the two targets, i.e. (*Pwt* + *Twt*) and (*Pwt* + *Tmut*).

$$\text{Case2, depletion : } S_2 = \ln \left\{ 1 + \frac{c_{Tmut}}{c_{Twt}} \exp\left(\frac{\Delta\Delta G_{Pmut} + \Delta\Delta G_{Pwt}}{RT}\right) \right\} \quad (3)$$

This predicts an increase in *S* when mutant target is present, which should translate into an improvement of sensitivity towards mutation detection. Quantitatively, this is determined by  $\Delta\Delta G_{Pwt}$  which also sets a physico-chemical boundary on the depletion-induced improvement [27].

### 3.2. Effect of depletion on the SNV detection sensitivity

We experimentally determined the effect of wild-type target depletion on the sensitivity of detection of low-abundant SNVs using the reduced probe sets selected in Section 2.3. To achieve this, two microarrays were designed to allow the simultaneous detection of the *EGFR* and *KRAS* gene fragments, with one array designed to induce depletion of the *KRAS* wt sequence and not of *EGFR* wt, and vice versa for the second array (Fig. 4).

For each of the two genes of interest we prepared a dilution series of the mutant sequence in a background of wild-type fragments. The concentration of target sequences optimal for microarray detection was found to be different between the depletion and the non-depletion case (see Supporting Information S3), which reflects the large difference in

the respective *Pwt* concentration on the arrays. To enable a comparison of the two conditions while avoiding artefacts due to either weak or saturated signals on the array, the analysis was carried out at a total target concentration of 0.01 nM for the non-depletion case and 1 nM for the depletion case.

In Fig. 5, the detection signal  $S - S_0$  is plotted, derived from the ratios of array spot intensities for the tested concentrations. *S* was defined in Eq. 1 to yield a baseline value  $S_0 \approx 0$  in the absence of mutant DNA. In the present experiment, however, the observed  $S_0$  differed between the depletion and non-depletion conditions (Supporting Information S4, Fig. S4). It is possible that the large difference in target concentration between the non-depletion and depletion conditions reveals second-order concentration-dependent effects in the behavior of *Pref*, see Supporting Information S5. This non-zero baseline is easily accounted for in the interpretation of the depletion-related sensitivity improvement by including  $S_0$  as an offset term in Eqs. (2) and (3):

$$\text{Case1, no depletion : } S_1 = S_0 + \ln \left\{ 1 + \frac{c_{Tmut}}{c_{Twt}} \exp\left(\frac{\Delta\Delta G_{Pmut}}{RT}\right) \right\} \quad (4)$$

$$\text{Case2, depletion : } S_2 = S_0 + \ln \left\{ 1 + \frac{c_{Tmut}}{c_{Twt}} \exp\left(\frac{\Delta\Delta G_{Pmut} + \Delta\Delta G_{Pwt}}{RT}\right) \right\} \quad (5)$$

For fitting of our experimental results, we rearranged these expressions as:

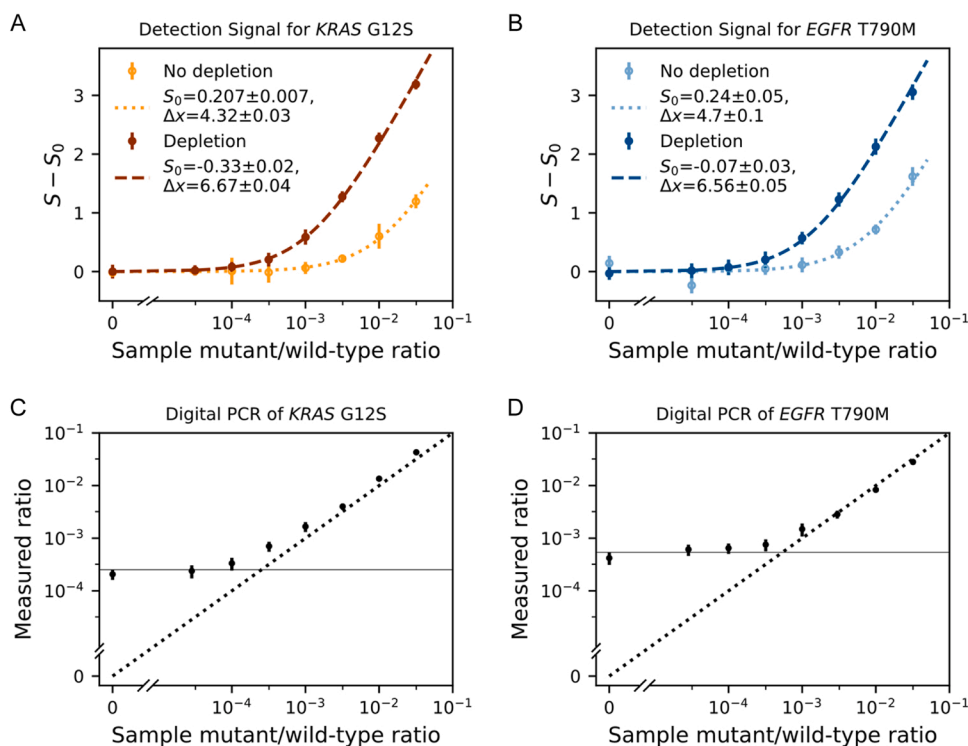
$$S = S_0 + \ln (1 + e^{x + \Delta x}) \quad (6)$$

where  $x = \ln\left(\frac{c_{Tmut}}{c_{Twt}}\right)$  is the logarithmic mutant ratio, which is a control parameter in our experiments and  $\Delta x = \Delta\Delta G/RT$ . A change in  $\Delta x$  corresponds to a horizontal shift in the signal curve and will directly impact the detection limit. Since we expect  $\Delta x_{non\ dep} = \Delta\Delta G_{Pmut}/RT$  in the non-depletion case and  $\Delta x_{dep} = (\Delta\Delta G_{Pmut} + \Delta\Delta G_{Pwt})/RT$  in the depletion case, the detection limit will differ between these two cases. Fitting the measured microarray data to this model, we obtained estimates for  $S_0$  and  $\Delta x$  shown in Fig. 5A and B. The expected  $\Delta x$  value can also be independently calculated from free energy parameters of our probe sequence. Using previously published  $\Delta\Delta G$  values for probe-target pairs containing mismatches, we find good agreement between these predicted  $\Delta x$  values and those derived from fitting our data, as shown in Table 2 [24].

We define the limit of detection (LoD) as the mutant fractional abundance corresponding to a signal of  $S_0 + 3 \times StdDev(S_0)$ . Calculating this value and applying it to the fitted dose-response functions, we obtain for *KRAS* G12S a limit of 0.38% in the non-depletion case and 0.054% in the depletion case (7-fold improvement), while for *EGFR* T790M the limits are 0.60% and 0.048%, respectively (12.5-fold improvement). Indeed, in both cases a ~10-fold increase in sensitivity is expected if the thermodynamically predicted values for  $\Delta\Delta G_{Pmut}$  and  $\Delta\Delta G_{Pwt}$  from Table 2 are applied to Eqs. (4) and (5).

### 3.3. Benchmarking SNV detection using dPCR

Digital PCR is considered a gold standard technique for low-abundant mutation detection. For comparison, we carried out dPCR measurements of the same sample dilution series, yielding an estimated LoD of 0.01% for *KRAS* G12S and 0.032% for *EGFR* T790M (Fig. 5C and D). Clinically validated methods for lung cancer mutation detection from liquid biopsy include next generation sequencing (NGS) (with a sensitivity of 99.48% for variant allele fractions of 0.25%) [36] and dPCR (for variant allele fractions of 0.5%) [37]. Highly specialized PCR methods such as BEAMing have shown to reach LoDs down to 0.01% but are complicated and time consuming; and therefore not easily implemented in clinical settings [38,39]. Thus, the depletion-driven



**Fig. 5.** Top: Detection signal ( $S - S_0$ , unitless) over a range of mutant ratios (unitless) for the *KRAS* G12S (A) and *EGFR* T790M (B) SNVs, with and without depletion. Lines represent a least squares fit to Eqs. (2) and (3) where  $S_0$  represents the baseline signal and  $\Delta x$  corresponds to a horizontal shift in the signal curve. Error bars represent the standard deviation of 20 replicate spots, while fit uncertainties quoted in the insets are the standard errors determined from the covariance matrix of the fit. Bottom: Digital PCR results carried out with the same mutant dilution as for the microarray experiments, for the *KRAS* G12S (C) and *EGFR* T790M (D) targets. The log-log plot shows the mutant fractional abundance measured via dPCR vs. the known input mutant ratios. The horizontal line indicates the 0.95 confidence interval of the 0% mutant sample. The dotted diagonal line indicates the expected 1:1 relation between the input and output.

**Table 2**

Calculated  $\Delta\Delta G_{Pmut}/RT$  and  $\Delta\Delta G_{Pwt}/RT$  values based on the probe sequences using thermodynamic nearest-neighbor parameters from Hadiwikarta et al. [24], which are used to derive a predicted  $\Delta x$  for the non-depletion and depletion cases. Corresponding fitted parameters for the signal curves, reproduced from Fig. 5A and B, are shown as  $\Delta x$  (fitted).

SNV	$\frac{\Delta\Delta G_{Pmut}}{RT}$	$\frac{\Delta\Delta G_{Pwt}}{RT}$	Non-depletion		Depletion	
			$\Delta x$ (predicted)	$\Delta x$ (fitted)	$\Delta x$ (predicted)	$\Delta x$ (fitted)
<i>KRAS</i> G12S	4.46	2.40	4.46	$4.32 \pm 0.03$	6.86	$6.67 \pm 0.04$
<i>EGFR</i> T790M	4.82	2.30	4.82	$4.7 \pm 0.1$	7.12	$6.56 \pm 0.05$

sensitivity improvement demonstrated here elevates performance to a level comparable with currently used clinical methods and suitable for liquid biopsy applications.

This approach can be further multiplexed. While in the present design depletion probes occupy most of the array area, these probes could instead be incorporated on the surface of micro-beads included in the hybridization volume, or the lid and walls of the hybridization vessel. Combining this with the fact that only three different probes are required for relative quantification of one SNV, the array area can be functionalized with multiple probe sets (one per SNV) which enables highly multiplexed detection. Further, the treatment presented here is not technology-dependent, and could be used to achieve sensitivity improvements across a wide range of hybridization-based sensing technologies regardless of readout modality, such as fiber-optic [40], electrochemical [41] or piezoelectric DNA biosensors [42].

#### 4. Conclusion

Liquid biopsies, which require only a minimally-invasive blood draw for analysis are an attractive sample format for diagnosis and monitoring of multiple diseases. Due to the often very low concentrations of target biomarkers in such samples, it is still a challenge to develop methods with sufficient sensitivity for this application.

This work builds on theoretical predictions of the effect of wild-type analyte depletion on DNA mutation sensing sensitivity. We demonstrate experimentally that, by promoting depletion of wild-type target

sequences, the limit of detection for an SNV of interest can be lowered by approximately an order of magnitude. Combining this depletion effect with a theoretical treatment based on physicochemical principles of DNA hybridization thermodynamics, we detect SNVs down to 0.05% fractional abundance in a wild-type background. This sensitivity is comparable with current benchmark methods, and suitable for clinical applications such as liquid biopsy-based gene analysis.

While we employ microarray technology in this study, the general approach being demonstrated is not dependent on a particular detection method and should be applicable to other hybridization-based DNA biosensing technologies. Therefore, this method has the potential to become a useful tool for enhancing the performance of various biosensing techniques and play a role in their implementation in emerging analytical fields such as liquid biopsy analysis.

#### Funding

This work was supported by the Research Foundation Flanders (FWO) and the Flemish Institute for Technological Research VITO nv (grant numbers 1S69320N, 1159719N) and by the EU H2020 M3DLoC project (grant number 760662).

#### CRedit authorship contribution statement

Rebekka Van Hoof and Michal Szymonik performed the majority of experimental design and execution, data analysis and manuscript

preparation. Stefanos K. Nomidis and Jef Hooyberghs provided the theoretical explanations. Karen Hollanders prepared all PCR products and An Jacobs performed the dPCR experiment. Jef Hooyberghs oversaw the design of the experiments and theoretical calculations. Inge Nelissen and Patrick Wagner contributed to the interpretation of the results and edited the manuscript. All authors read and approved the final manuscript.

## Notes

The authors declare that there is no conflict of interest.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.snb.2022.132175](https://doi.org/10.1016/j.snb.2022.132175).

## References

- Saifullah, S. Fuke, H. Nagasawa, T. Tsukahara, Single nucleotide recognition using a probes-on-carrier DNA chip, *Biotechniques* 66 (2019) 73–78, <https://doi.org/10.2144/btn-2018-0088>.
- H. Willems, A. Jacobs, W.W. Hadiwikarta, T. Venken, D. Valkenburg, N. Van Roy, et al., Thermodynamic framework to assess low abundance DNA mutation detection by hybridization, *PLoS One* 12 (2017), e0177384, <https://doi.org/10.1371/journal.pone.0177384>.
- B. van Grinsven, N. Vanden Bon, L. Grieten, M. Murib, S.D. Janssens, K. Haenen, et al., Rapid assessment of the stability of DNA duplexes by impedimetric real-time monitoring of chemically induced denaturation, *Lab Chip* 11 (2011) 1656–1663, <https://doi.org/10.1039/C1LC20027E>.
- B. van Grinsven, N. Vanden Bon, H. Strauven, L. Grieten, M. Murib, K.L. Jiménez Monroy, et al., Heat-transfer resistance at solid–liquid interfaces: a tool for the detection of single-nucleotide polymorphisms in DNA, *ACS Nano* 6 (2012) 2712–2721, <https://doi.org/10.1021/nn300147e>.
- K. Knez, D. Spasic, K.P.F. Janssen, J. Lammertyn, Emerging technologies for hybridization based single nucleotide polymorphism detection, *Analyst* 139 (2014) 353–370, <https://doi.org/10.1039/C3AN01436C>.
- S. Galbiati, F. Damin, V. Burgio, A. Brischi, N. Soriani, B. Belcastro, et al., Evaluation of three advanced methodologies, COLD-PCR, microarray and ddPCR, for identifying the mutational status by liquid biopsies in metastatic colorectal cancer patients, *Clin. Chim. Acta* 489 (2019) 136–143, <https://doi.org/10.1016/j.cca.2018.12.004>.
- W.W. Hadiwikarta, B. Van Dorst, K. Hollanders, L. Stuyver, E. Carlon, J. Hooyberghs, Targeted resequencing of HIV variants by microarray thermodynamics, *Nucleic Acids Res* 41 (2013), e173, <https://doi.org/10.1093/nar/gkt682>.
- D. Khodakov, C. Wang, D.Y. Zhang, Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches, *Adv. Drug Deliv. Rev.* 105 (2016) 3–19, <https://doi.org/10.1016/j.addr.2016.04.005>.
- M. Stancescu, T.A. Fedotova, J. Hooyberghs, A. Balaëff, D.M. Kolpashchikov, Nonequilibrium hybridization enables discrimination of a point mutation within 5–40 °C, *J. Am. Chem. Soc.* 138 (2016) 13465–13468, <https://doi.org/10.1021/jacs.6b05628>.
- A. Kowalczyk, Trends and perspectives in DNA biosensors as diagnostic devices, *Curr. Opin. Electrochem.* 23 (2020) 36–41, <https://doi.org/10.1016/j.coelec.2020.03.003>.
- K. V, DNA biosensors-A review, *Journal of Bioengineering & Biomedical Science* 07, 2017. <https://doi.org/10.4172/2155-9538.1000222>.
- Y. Hua, J. Ma, D. Li, R. Wang, DNA-based biosensors for the biochemical analysis: A review, *Biosensors* 12 (2022) 183.
- J.P. Tosar, G. Brañas, J. Laif, Electrochemical DNA hybridization sensors applied to real and complex biological samples, *Biosens. Bioelectron.* 26 (2010) 1205–1217, <https://doi.org/10.1016/j.bios.2010.08.053>.
- V.V. Demidov, M.D. Frank-Kamenetskii, Two sides of the coin: affinity and specificity of nucleic acid interactions, *Trends Biochem. Sci.* 29 (2004) 62–71, <https://doi.org/10.1016/j.tibs.2003.12.007>.
- I. Ladas, M. Fitarelli-Kiehl, C. Song, V.A. Adalsteinsson, H.A. Parsons, N.U. Lin, et al., Multiplexed elimination of wild-type DNA and high-resolution melting prior to targeted resequencing of liquid biopsies, *Clin. Chem.* 63 (2017) 1605–1613, <https://doi.org/10.1373/clinchem.2017.272849>.
- A. Abi, A. Safavi, Targeted detection of single-nucleotide variations: Progress and promise, *ACS Sens.* 4 (2019) 792–807, <https://doi.org/10.1021/acssensors.8b01604>.
- T. Seyama, T. Ito, T. Hayashi, T. Mizuno, N. Nakamura, M. Akiyama, A novel blocker-PCR method for detection of rare mutant alleles in the presence of an excess amount of normal DNA, *Nucleic Acids Res.* 20 (1992) 2493–2496, <https://doi.org/10.1093/nar/20.10.2493>.
- S.T. Lee, J.Y. Kim, M.J. Kwon, S.W. Kim, J.H. Chung, M.J. Ahn, et al., Mutant enrichment with 3'-modified oligonucleotides a practical PCR method for detecting trace mutant DNAs, *J. Mol. Diagn.* 13 (2011) 657–668, <https://doi.org/10.1016/j.jmoldx.2011.07.003>.
- P.L. Dominguez, M.S. Kolodney, Wild-type blocking polymerase chain reaction for detection of single nucleotide minority mutations from clinical specimens, *Oncogene* 24 (2005) 6830–6834, <https://doi.org/10.1038/sj.onc.1208832>.
- X. Sun, K. Hung, L. Wu, D. Sidransky, B. Guo, Detection of tumor mutations in the presence of excess amounts of normal DNA, *Nat. Biotechnol.* 20 (2002) 186–189, <https://doi.org/10.1038/nbt0202-186>.
- J.S. Wang, D.Y. Zhang, Simulation-guided DNA probe design for consistently ultraspecific hybridization, *Nat. Chem.* 7 (2015) 545–553, <https://doi.org/10.1038/nchem.2266>.
- T. Wu, W. Chen, Z. Yang, H. Tan, J. Wang, X. Xiao, et al., DNA terminal structure-mediated enzymatic reaction for ultra-sensitive discrimination of single nucleotide variations in circulating cell-free DNA, *Nucleic Acids Res.* 46 (2017), e24, <https://doi.org/10.1093/nar/gkx1218>.
- J. Hooyberghs, M. Baiesi, A. Ferrantini, E. Carlon, Breakdown of thermodynamic equilibrium for DNA hybridization in microarrays, *Phys. Rev. E* 81 (2010), 012901, <https://doi.org/10.1103/PhysRevE.81.012901>.
- W.W. Hadiwikarta, J.C. Walter, J. Hooyberghs, E. Carlon, Probing hybridization parameters from microarray experiments: nearest-neighbor model and beyond, *Nucleic Acids Res* 40 (2012), e138, <https://doi.org/10.1093/nar/gks475>.
- W.W. Hadiwikarta, E. Carlon, J. Hooyberghs, Dynamic range extension of hybridization sensors, *Biosens. Bioelectron.* 64 (2015) 411–415, <https://doi.org/10.1016/j.bios.2014.09.043>.
- H.N. Wood, T. Venken, H. Willems, A. Jacobs, A.J. Reis, P.E. Almeida da Silva, et al., Molecular drug susceptibility testing and strain typing of tuberculosis by DNA hybridization, *PLoS One* 14 (2019), e0212064, <https://doi.org/10.1371/journal.pone.0212064>.
- S.K. Nomidis, M. Szymonik, T. Venken, E. Carlon, J. Hooyberghs, Enhancing the performance of DNA surface-hybridization biosensors through target depletion, *Langmuir* 35 (2019) 12276–12283, <https://doi.org/10.1021/acs.langmuir.9b01761>.
- R. Nussinov, H. Jang, C.-J. Tsai, F. Cheng, Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers, *PLoS Comput. Biol.* 15 (2019), e1006658, <https://doi.org/10.1371/journal.pcbi.1006658>.
- C. Abbosh, N.J. Birkbak, G.A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, et al., Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution, *Nature* 545 (2017) 446–451, <https://doi.org/10.1038/nature22364>.
- X. Liu, J. Lang, S. Li, Y. Wang, L. Peng, W. Wang, et al., Fragment enrichment of circulating tumor DNA with low-frequency mutations, *Front. Genet.* 11 (2020) 147, <https://doi.org/10.3389/fgene.2020.00147>.
- L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, Global cancer statistics, 2012, *CA: Cancer J. Clin.* 65 (2015) 87–108, <https://doi.org/10.3322/caac.21262>.
- C.J. Langer, Roles of EGFR and KRAS mutations in the treatment of patients with non-small-cell lung cancer, *P T* 36 (2011) 263–279.
- S. Institute, COSMIC v94, 2021. (<https://cancer.sanger.ac.uk/cosmic>) (accessed 16 November 2021).
- EMBL-EBL LRG: Stable reference sequences for reporting variants, 2021. (<http://www.lrg-sequence.org/>) (accessed 16 November 2021).
- H.R. Underhill, J.O. Kitzman, S. Hellwig, N.C. Welker, R. Daza, D.N. Baker, et al., Fragment length of circulating tumor DNA, *PLoS Genet* 12 (2016), e1006162, <https://doi.org/10.1371/journal.pgen.1006162>.
- V. Plagnol, S. Woodhouse, K. Howarth, S. Lensing, M. Smith, M. Epstein, et al., Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling, *PLoS One* 13 (2018), e0193802, <https://doi.org/10.1371/journal.pone.0193802>.
- C. Silveira, A.C. Sousa, A. Janeiro, S. Malveiro, E. Teixeira, E. Brysch, et al., Detection and quantification of EGFR T790M mutation in liquid biopsies by droplet digital PCR, *Transl. Lung Cancer Res* 10 (2021) 1200–1208, <https://doi.org/10.21037/tlcr.20-1010>.
- K. Taniguchi, J. Uchida, K. Nishino, T. Kumagai, T. Okuyama, J. Okami, et al., Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas, *Clin. Cancer Res.* 17 (2011) 7808–7815, <https://doi.org/10.1158/1078-0432.CCR-11-1712>.
- J. Zhao, H.H. Feng, J.Y. Zhao, L.C. Liu, F.F. Xie, Y. Xu, et al., A sensitive and practical method to detect the T790M mutation in the epidermal growth factor receptor, *Oncol. Lett.* 11 (2016) 2573–2579, <https://doi.org/10.3892/ol.2016.4263>.

- [40] G. Jiang, R. Chen, H. Yan, Q. Ouyang, A new method of preparing fiber-optic DNA biosensor and its array for gene detection, *Sci. China Life Sci.* 44 (2001) 33–39, <https://doi.org/10.1007/BF02882070>.
- [41] M. Trotter, N. Borst, R. Thewes, F. von Stetten, Review: Electrochemical DNA sensing – Principles, commercial systems, and applications, *Biosens. Bioelectron.* 154 (2020), 112069, <https://doi.org/10.1016/j.bios.2020.112069>.
- [42] M. Pohanka, Overview of piezoelectric biosensors, immunosensors and DNA sensors and their applications, *Materials* 11 (2018) 448, <https://doi.org/10.3390/ma11030448>.

Rebekka Van Hoof obtained her master's degree in Biochemistry and Biotechnology at the KU Leuven. Currently she is a PhD candidate at the VITO Health Diagnostics group in collaboration with the Laboratory for Soft Matter and Biophysics at the KU Leuven and the Theory Lab and the Data Science Institute at UHasselt. She researches the development of technologies and sensors for the targeted detection of nucleic acid mutation derived from extracellular vesicles, a biomarker in liquid biopsies.

Michał Szymonik received a PhD from the School of Electrical and Electronic Engineering at the University of Leeds, England in 2012. He currently works at the Flemish Institute for Technological Research (VITO) in Belgium, participating in a various European initiatives in the healthcare technology space. His main research interest is the development of point-of-care biosensor technologies for medical diagnostics, incorporating various sensing modalities and biorecognition elements.

Stefanos K. Nomidis studied Applied Physics and Mathematics at the National Technical University of Athens (NTUA). He then relocated to Belgium, where he obtained his master's and doctoral degree in Physics at the KU Leuven. His research focused on theory and simulations of DNA mechanics and hybridization. He is currently working as an AI Engineer at Relu, a startup company that develops AI-assisted segmentation software for Dental and Maxillofacial Images.

Karen Hollanders is trained as lab technician in Pharmaceutical and Biological Laboratory Technology (Leuven, 2001). She currently holds the position of lab technician in the VITO Health group.

An Jacobs is trained as lab technician in biochemistry and pharmacology. She obtained her bachelor's degree in 1999 and is a senior lab technician in the VITO Health Diagnostics group since 2009.

Inge Nelissen is trained as Industrial Engineer in Biochemistry, Bio-engineer in Biotechnology, and obtained a PhD in Medical Sciences at the KU Leuven in 2003. She currently holds the position of Project Manager and Team Leader in the VITO Health Diagnostics group. Her team focuses on development and validation of technology for biomarker enrichment and detection in liquid biopsies.

Patrick Wagner received his PhD in Physics in 1994 at TU Darmstadt and was postdoctoral researcher at KU Leuven from 1995 till 2001. In 2001, he was appointed as a professor for biophysics at UHasselt and he returned in 2014 as full professor back to KU Leuven. His main research field is the development of bioanalytical sensors for diagnostic and environmental applications with a special focus on synthetic receptors and thermal detection principles. He has authored or co-authored 300 publications and received several scientific distinctions. Since 2016, P. Wagner is Editor-in-Chief of the new Elsevier journal *Physics-in-Medicine*.

Jef Hooyberghs obtained a master and a PhD in science, physics at the KU Leuven and UHasselt. Currently, he is affiliated to VITO, the Flemish research organization in the area of cleantech and sustainable development, and part time professor at UHasselt. At VITO he is research leader and head of the Data Science team, at UHasselt he is member of the Theory Lab and the Data Science Institute.