Sensitivity enhancement in lateral flow assays: a systems perspective

Joshua D. Bishop, Helen V. Hsieh, David J. Gasperino and Bernhard H. Weigl*

Lateral flow assays (LFAs) are rapid, inexpensive, easy-to-manufacture and -use tests widely employed in medical and environmental applications, particularly in low resource settings. Historically, LFAs have been stigmatized as having limited sensitivity. However, as their global usage expands, extensive research has demonstrated that it is possible to substantially improve LFA sensitivity without sacrificing their advantages. In this critical review, we have compiled state-of-the-art approaches to LFA sensitivity enhancement. Moreover, we have organized and evaluated these approaches from a system-level perspective, as we have observed that the advantages and disadvantages of each approach have arisen from the integrated and tightly interconnected chemical, physical, and optical properties of LFAs.

1. Introduction

Lateral flow assays (LFAs) have been developed for use in a range of fields, such as medical diagnostics, veterinary medicine, food safety, and environmental safety, to serve as analytical tests for a range of biochemical analytes, such as anti-
and assay reagents are transported by capillary action. A standard LFA test strip accepts sample (and a running buffer, if required) on a sample pad, stores detection reagents (i.e., conjugates) in a conjugate pad, arranges immobilized capture reagents in the form of test and control lines on an analytical membrane, and draws excess fluid into an absorbent pad. The test strip may be enclosed in a cassette or run as a dipstick in a test tube. The result of an LFA is almost always related to an optical signal generated at the test line. In a sandwich assay format, a signal is generated by the concentration of detection reagent bound to captured analyte. Alternatively, in a competitive assay format (commonly used for small targets such as drugs of abuse), a signal at the test line is eliminated by the displacement of captured detection reagent with the analyte.

The ideal LFA is sensitive, specific, stable, fast (<30 minutes), inexpensive, and user-friendly. It gives reproducible results and is robust to variation in patient samples and operating parameters such as sample and running buffer volumes. However, these performance characteristics arise from the integrated and interconnected chemical, physical, and optical properties of the LFA (Fig. 2). The complexity of these interconnections suggests that a system-level understanding will be critical to the realization of the ideal LFA and the delineation of its fundamental weaknesses.

In this review, we have categorized improvements in LFA analytical sensitivity from a high-level systems perspective using three organizing principles—reaction (section 2), transport (section 3), and signal (section 4). Furthermore, we have categorized the transformative improvements that connect at least two of these organizing principles (section 5) and noted how these interconnected improvements invite a systems approach to LFA development. We have constrained the scope of this review, by use of these organizing principles, to be within the operational bounds of the LFA. Therefore, we have not discussed improvements to sensitivity that arise from innovations in reagent or material manufacturing processes, or non-integrated sample preparation steps, for example.

All the same, the variety of approaches that fit within our framework is quite large. The magnitude of recent literature on this topic speaks not only to the importance of the LFA and its broad applicability, but also to the complexity of the conventional LFA, in that it can support such a wide range of modifications. Due to the breadth of that range, which can include differences in test format, analytical target, read-out method, instrumentation, etc., we chose to provide—where reported—the fold-change difference observed in relation to an unmodified, conventional LFA system.

In addition to our categorization of recent advances in LFA sensitivity enhancement, we have endeavoured to critically assess the limitations on approaches within each of our categories. We have found it often to be the case that innovations designed to improve sensitivity in LFAs have degraded...
other important performance characteristics (e.g. test time, user steps, test autonomy) and/or have not been tested in a fully implemented test format against a representative set of clinical or environmental samples. Although the focus of our review is on sensitivity enhancements, it is these general observations that back our advocacy for a system-level approach to LFA development and innovation.

2. Reaction

The reactions that occur within an LFA strip are most often in the same format as a "sandwich" ELISA. In this format, an immobilized capture reagent (typically an antibody) binds an analyte (typically a protein antigen with multiple epitopes) to which binds a labelled detection reagent (typically one or more different antibodies conjugated to a nanoparticle label). Ideally, either a significant amount of detection reagent or none is localized to the capture reagent at equilibrium, depending on analyte presence or absence, respectively.

Perhaps the most obvious way to improve the performance of the LFA reaction with respect to these conditions is to introduce capture and detection reagents with high affinity and/or avidity for the analyte. Less obvious is what the best approach to obtaining or creating new, high-affinity reagents should be. Examples of demonstrated approaches include the adaptation of naturally occurring high-affinity reagents and the design of synthetic, de novo high-affinity reagents.

Natural

Adaptations of streptavidin and biotin as assay reagents in the LFA format take advantage of a binding pair that is one of the strongest found in nature, is highly resistant to biochemical interference, and is naturally multivalent in that wild-type streptavidin binds four biotins. BioProto has commercialized the generic RapidAssay Device (gRAD), in which analyte binds to a labelled detection reagent and a non-immobilized, biotinylated capture reagent upstream of the test line. This complex then binds to an immobilized, biotin-binding reagent at the test line. Yang et al. demonstrated a fast (10 minutes) gRAD for foot-and-mouth disease with ELISA-like sensitivity.8 Similarly, Expedeon Inc. (San Diego, CA, USA) has launched a Universal Lateral Flow Assay Kit, in which the non-immobilized capture reagent contains a "ULFA-Tag" and the LFA test line is made of immobilized anti-ULFA-Tag antibody.

Synthetic

The creation of de novo binding reagents takes advantage of increasingly powerful computational models that can target specific epitopes, search over multiple interdependent reagent characteristics (e.g. size, stability, charge), and introduce multivalent binding. For example, Strauch et al. demonstrated the use of two computationally designed hemagglutinin-binding proteins, specifically designed to bind different analyte epitopes, as detection and capture probes in an influenza LFA.9 However, these reagents were not reported as tested with clinical influenza samples, or against an existing influenza LFA for the purpose of demonstrating improved sensitivity.

Limitations

The reaction category covers the high-affinity interactions between analyte and the capture and/or detection reagents, whether implemented in a sandwich or a competitive assay. In seeking to introduce the highest possible affinity reagents to conventional LFA systems, the two approaches categorized here ultimately suffer from a problem of scale. Currently, the number of natural high-affinity reagents commonly used in an LFA is limited to a single affinity pair (streptavidin–biotin) that can only be used in one place within an LFA reaction. Conversely, the number de novo proteins that are candidates for use as high-affinity reagents in LFAs is unknown, but likely astronomical. Unfortunately, while the tools for designing, producing, and screening de novo proteins are continuously improving, the development effort behind even a single de novo reagent is currently considerable. In addition, this extensive development process is primarily underwritten by therapeutic applications, which pose a set of design specifications that are not necessarily compatible with diagnostic applications. Nanobodies are examples of therapeutic agents with obvious diagnostic applications,10 and they are easier to produce, smaller, and more stable than antibodies, but they still require extensive optimization and screening similar to that for antibodies. Aptamers are examples of "programmable," nucleic-acid-based binding reagents that have shown a
range of sensitivities from mid-picomolar to high-micromolar, which argues a continued need for case-by-case optimization even though—or especially because—they introduce new reaction sets to the LFA format. Ultimately, the production process of any LFA reagent—antibody, de novo protein, or otherwise—must align with the most desirable characteristics of the LFA itself: integrated, low-cost, and rapid. Without this alignment the innovations in this category are unlikely to make an impact on the practice of LFA development.

3. Transport

The transport of sample, buffers, and assay reagents in an LFA strip is provided by the capillary action that occurs within its constituent porous materials. Reaction and transport are inextricably linked, in that changes to transport characteristics such as volumes or flow rates necessarily change reaction kinetics. Many fluidic controls have been introduced to tune transport dynamics, and therefore to tune reaction kinetics, and have been demonstrated to improve sensitivity in paper-based systems. Examples include modified sample and conjugate pad geometries (8-fold sensitivity improvement in a human IgG LFA), modified flow rates using a sponge shunt (10-fold sensitivity improvement in a nucleic acid LFA), and modified flow rates using wax pillars (3-fold sensitivity improvement in a human IgG LFA). Many other paper-based fluidic controls were recently reviewed.

However, one area in which physical, transport-based approaches to sensitivity improvement can be largely distinguished from biochemical, reaction-based approaches is sample preparation. Several on-board sample preparation methods that selectively transport analyte—through the physical separation and/or concentration of analyte from sample—to improve sensitivity have been demonstrated on LFAs. For example, the size-based separation of sample components by filtration is easily achieved in a system of porous materials, like an LFA. Commercial vendors, such as GE Healthcare Life Sciences (Pittsburgh, PA USA) and Pall Life Sciences (Ann Arbor, MI USA), offer plasma separation membranes that can be integrated into the LFA strip format without affecting the standard LFA workflow.

Other selective transport methods that have recently been applied to LFAs include separation and/or concentration of analyte from sample by electrophoresis, aqueous two-phase systems, and dialysis membranes.

Electrophoretic

Electrophoretic methods drive transport of charged particles with applied electric fields. Demonstrated electrophoretic sample separation methods in paper-based devices include the direct application of electric fields, ion concentration polarization (ICP), and isochromophoresis (ITP). Wu et al. applied a direct electric field to selectively enhance the transport of nucleic acids, and improve sensitivity 400-fold, in an LFA that detected PCR products from an H5N1 influenza PCR assay.

Although the application of ICP to micro/nanofluidic platforms was recently reviewed, there are some examples of ICP in paper-based devices. Gong et al. demonstrated ICP in a saturated paper-based device, which improved sensitivity 5-fold for a FITC–BSA conjugate and 4-fold for bromoresol green dye compared to without ICP; while Liu et al. demonstrated ICP in a Parafilm- and paper-based device, which concentrated FITC. Lastly, Moghadam et al. demonstrated ITP and improved sensitivity 400-fold in a goat anti-human IgG LFA.

Phasic

Aqueous two-phase systems (ATPS) use different water-soluble materials (e.g. polymers, salts, or micelles) at high concentration to form two immiscible aqueous phases, allowing concentration of a target protein in one phase. While ATPS were demonstrated to work as a pre-concentration step before the application of sample to a bacteriophage M13 and a transferrin LFA, Pereira et al. integrated the APTS step into a “three-dimensional” design to improve sensitivity 10-fold in a Plasmodium lactate dehydrogenase (pLDH) LFA.

Dialytic

Dialysis is a common technique applied to concentrate nucleic acids and proteins, and to deplete a sample of small contaminants. Tang et al. demonstrated nucleic acid dialysis and protein dialysis in separate, one-step LFA devices. They integrated a dialysis membrane and PEG-based water absorption onto LFA strips, and concentrated (separately) HIV nucleic acid and myoglobin analytes, which improved sensitivity 10-fold and 4-fold in the respective LFAs.

Limitations

The transport category contains an important set of optimization parameters for an LFA. Transport dynamics in an LFA should be tuned during development to optimize the integrated binding reactions (specific and non-specific) and optical signal generation. There are several design elements that can be changed to tune LFA transport dynamics, including the type, geometry, and layout of porous materials, but also the constitution of various buffers and the physical and chemical characteristics of reagents. Selective transport, on the other hand, represents a set of sensitivity enhancement approaches in which initial sample preparation steps are prepended to the standard LFA workflow. Here, as in the more interconnected reaction/transport and reaction/transport/optical approaches discussed below, more sophisticated transport systems support more sophisticated workflows, but while the additional sophistication might lead to improved sensitivity, it is often at the cost of added test time, user steps, and/or instrumentation. Future innovations in this category should look to minimize these trade-offs.
4. Signal

The signal coming from an LFA test strip is most often a change in optical intensity generated by the analyte-associated aggregation (or disaggregation, in a competitive format) of detection reagents on the test line. In the simplest case, such as when a coloured label is part of the detection reagent, the signal can be interpreted as colorimetric signals, directly by eye of the operator under ambient conditions. The earliest and most commonly used coloured labels are colloidal gold (typically 20–40 nm) and coloured latex labels. These simple, spherical labels do not directly bind analyte and must be conjugated to an analyte-binding reagent to form the detection reagent. Despite the ubiquity of these basic labels, many other types of labels, some of which generate signal by alternate mechanisms, have improved sensitivity in LFAs.

Colorimetric labels

For example, various shapes of gold labels have been explored. While alternative gold label shapes—including nanorods, nanowires, nanoshells, and nanocubes—are commercially available from vendors such as Nanopartz (Loveland, CO USA) and Sono Nanotech (Halifax, Nova Scotia Canada), reports of their use as labels in LFAs are limited. However, a study comparing thermal contrast and optical contrast detection methods indicated that gold nanorods and nanoshells may improve sensitivity of LFAs compared to standard colloidal gold nanoparticles. Likewise, gold nanorod labels were investigated during development of a competitive LFA for breast cancer testing. The developers characterized mono-dispersed gold nanorods with various aspect ratios, and tested methods of bio-functionalization (to create detection reagents) and immobilization (to create capture reagents). Separately, Yang et al. reported that a proof-of-concept IgG LFA with gold nanocage labels yielded 2.5-fold improved sensitivity over the same system with colloidal gold labels.

Multimeric gold labels can also improve assay sensitivity. For example, Shen and co-authors used dendrimers to assemble gold label multimers, which improved their assay sensitivity by 20-fold. Hu et al. used complementary oligonucleotides to create gold label multimers. Their analysis found that the colour intensity of the multimers was stronger than either monomer, which led to 2.5-fold improved sensitivity in their HIV nucleic acid LFA. Micron-sized labels, created from long (3–7 µm) silica nanorods decorated with smaller (10 nm) gold nanoparticles, improved sensitivity 50-fold in a rabbit IgG LFA. Similarly, a large multi-mer material label, composed of many small (16 nm) gold labels immobilized to individual chitosan-coated Fe$_3$O$_4$ nanoparticles (chosen to enhance reagent preparation through magnetic separation), improved sensitivity 3-fold for a competitive aflatoxin B$_2$ assay in a qualitative LFA.

A variety of other colorimetric labels have improved sensitivity compared to gold or latex labels. Dye-filled liposomes can improve sensitivity compared to gold and latex labels, due to the large amount of dye molecules they can contain, but they can be difficult to stabilize. NanoAct labels (Asahi Kasei, Miyazaki, Japan) are cellulose nanobeads that were reported to yield 10-fold higher sensitivity than gold or latex labels in an influenza LFA. The authors attribute the improved sensitivity to the larger surface area of the NanoAct labels (330 nm), and more intense dye colours available for impregnation into the cellulosic material. Carbon nanoparticle labels were reported to improve sensitivity 100-fold over gold label in a dengue antigen NS1 LFA. They provide high visual contrast to the analytical membrane and are inexpensive to produce, but they are only available from a limited number of vendors. Comparatively, carbon nanotube labels form high surface area test lines, which were reported to improve sensitivity 5-fold in an LFA with DNA oligo detection and 10-fold in an hCG LFA.

Non-colorimetric labels

Yet other, non-colorimetric labels can generate signals other than a change in visible-by-eye optical intensity, such as fluorescent dyes, luminescent molecules, quantum dots, upconverting phosphors, and magnetic and SERS particles; these have been recently reviewed for improvement of sensitivity in LFAs. Advantages of these labels over colorimetric labels include enhanced signal range, and added objectivity and potential quantification provided by associated reader instrumentation; disadvantages include detection reagent complexity, and the cost and complexity of the reader.

Commercial readers are available for detection of colorimetric and non-colorimetric labels. They are sold as stand-alone units, for example from Qiagen Lake Constance (Stockach, Germany), LRE (Munich, Germany), Hamamatsu (Tokyo, Japan), and Axxin (Fairfield, Australia); or integrated into the LFA itself, for example the BD Veritor and Clearblue tests. In addition, to reduce reader-associated test costs, particularly for low-resource settings, recent research has been conducted on the use of smartphones as LFA readers.

Near-infrared (NIR) fluorescent dyes were reported to allow the sensitivity of IL-6 and C-reactive protein LFAs to match that of traditional ELISAs. Additionally, as these dyes absorb and emit fluorescence at wavelengths outside of typical biological matrices, these results were obtained for analytes spiked into 10% plasma matrices. Chen and Wu used a biotin-functionalized antibody, biotin-functionalized polylysine scaffolds loaded with multiple fluorescent dyes, and streptavidin linkages to create a detection reagent with tens to hundreds of fluorescent dye labels. These multivalent labels increased the assay sensitivity 100-fold compared to gold labels in a Cry1Ab toxin LFA. Silver nanoparticles were used to quench carbon dots in a zearalenone LFA that was ten times more sensitive than the LFA with silver nanoparticle labels alone.

Up-converting phosphor labels were reported to yield greater than 10-fold sensitivity in a Schistosoma LFA.
compared to a traditional ELISA,\textsuperscript{51} and equivalent sensitivity to a liquid chromatography tandem mass spectrometry test in morphine and methamphetamine LFAs.\textsuperscript{52} Luminescent labels were reported to return a 10-fold improvement in a lysosome LFA using a time-resolved reader;\textsuperscript{53} luminescent quantum dots have contributed two orders of magnitude improved sensitivity over colloidal gold in an LFA for aflatoxin in maize extract.\textsuperscript{54}

Non-optical labels

Non-optical signals from LFAs become measurable with readers added as instrumentation. Thermal contrast imaging readers optically stimulate metallic (such as gold) nanoparticle labels and measure the generated heat in areas of metallic label aggregation, such as the test line of an LFA. Measuring thermal contrast increased sensitivity 32-fold in an FDA-approved cryptococcal antigen LFA\textsuperscript{30} and 8-fold over visual or colorimetric readers for influenza, malaria, or Clostridium difficile.\textsuperscript{55} Advantages for thermal contrast readers include compatibility with already available commercial tests. Additionally, thermal contrast imaging readers with low-cost components have been designed for application in low-resource settings,\textsuperscript{55} reducing cost for implementation. Our FlowDX group at Intellectual Ventures laboratory has developed a field-rugged, stand-alone thermal reader with an onboard advanced algorithm for data analysis.\textsuperscript{56} Gold labels provide opportunities for detection of other signal modalities, including photoacoustic\textsuperscript{57} and X-ray\textsuperscript{58} imaging. The former technique was reported to yield a 100-fold improvement over colorimetric detection in a cryptococcal antigen LFA, while the latter technique has not been applied to the LFA format to our knowledge.

Other non-optical signals become measurable with readers and alternative labels. For example, the use of surface-enhanced Raman scattering (SERS) improved the sensitivity of a staphylococcal enterotoxin B LFA by four orders of magnitude compared to a commercial point-of-care test for the same analyte.\textsuperscript{59} Similarly, gold nanoparticle labels coated with 4-aminothiophenol enabled SERS detection in a neomycin and quinolones antibiotics LFA, with sensitivity improved three orders of magnitude over visual detection.\textsuperscript{60} Another group reported SERS detection of gold nanostars in a multiplexed LFA that distinguished between non-structural protein 1 from Zika and dengue viruses, with 15-fold and 7-fold improved sensitivity for the respective targets.\textsuperscript{13,61}

Magnetic labels can be detected with a magnetic reader, as demonstrated by Shi et al.\textsuperscript{62} with a Listeria monocytogenes LFA. These authors used superparamagnetic nanoparticle labels and a magnetic reader (Magna Biosciences, San Diego, CA) to generate a 10-fold improvement in assay sensitivity. A similar degree of sensitivity improvement was reported for a 3-amino-2-oxazolidinone (AOZ) LFA that used magnetic labels.\textsuperscript{63}

Catalytic labels are most often used for the amplification of colorimetric signals (which we have categorized in the reaction/signal section below), but Lin et al. showed that platinum label-mediated catalysis of hydrogen peroxide can produce a measurable change in pressure inside gas-tight containers. They reported comparable sensitivity, having tested 45 clinical blood samples for myoglobin, for their pressure-based LFA and a chemiluminescent microparticle immunoassay (CMIA), which is the clinical gold standard.\textsuperscript{64}

Limitations

The signal category covers signals that can be generated from a conventional LFA reaction, which is fundamentally limited to a proportional measure of the amount of analyte in a sample. As such, the sensitivity of a conventional LFA, even after enhancement by alternative labels and reader instrumentation, will never approach the sensitivity of a nucleic acid amplification test, for example, in which the signal is amplified exponentially. The possibility of enhancing the sensitivity of an LFA by several orders of magnitude is what has led to extensive research on integrating signal amplification reactions into the LFA format. We discuss these signal amplification approaches in the reaction/signal category below, although they have their own drawbacks due to their typically substantial deviations from the conventional LFA workflow. Those drawbacks keep the door open for innovations in label and reader technology that do not require deviations from the conventional LFA workflow, as they may be generically applied to current LFA technology. A high degree of compatibility with current LFA development and manufacturing practices are advantages for new labels and readers that may outweigh fundamental limitations on sensitivity improvements until more transformative, system-level innovations are commercially proven.

5. Transformative innovations occur in areas of overlap for reaction, transport, and signal

5.1 Reaction/transport

The integration of assay reactions and reagent transport enables LFAs to serve as an automated, single-step platform for running what are multi-step assays in other formats, such as plate-based ELISAs. The sensitivity of LFAs can be improved by carefully balancing reaction kinetics and transport dynamics. For example, screening the assay reactions on analytical membranes with various fluid flow rates is typically an early optimization step in a standard LFA development process. However, more transformative innovations go beyond the standard optimization efforts to explore non-standard arrangements of reaction and transport.

Reaction automation. The arrangement of materials in a standard LFA dictates an assay reaction order in which the detection reagent—also often called the “conjugate”—is stored dry in the conjugate pad; the detection reagent binds the analyte as the sample is transported from the sample pad through the conjugate pad on the way to the test line, where
the capture reagent is immobilized. This reaction order has been called premix delivery.

A non-standard order is sequential delivery, in which analyte is transported to the test line before (instead of with) the detection reagent. Sequential delivery requires a non-standard arrangement of materials to automate, because the detection reagent cannot simply be rehydrated by the added sample. However, it can be more sensitive than the premix delivery, as demonstrated by Liang et al. in a paper-based malaria assay.65 This demonstration included an analysis of the intensity across the test line (in the direction of flow) for premix compared to sequential delivery. Premix delivery led to a uniform distribution of labels on the test line. Sequential delivery led to a non-uniform distribution of labels on the test line, with more labels located upstream than the downstream. The authors make the argument, based on these distributions, that the kinetics of the reaction between capture reagents and the analyte/detection reagent complexes are slower than the kinetics of the reactions between the capture reagents and the analytes, and the captured analytes and the detection reagents.

These and other researchers in the Yager laboratory have demonstrated more non-standard reaction and transport dynamics in LFAs by moving from the strip format to a two-dimensional paper network (2DPN) format.66 For example, they modelled transport in 2DPNs67 and with sequential delivery.68 and produced 2DPNs with the sequential delivery of signal amplification reagents for 4-fold sensitivity improvement of a Plasmodium falciparum histidine-rich protein 2 (PfHRP2) LFA.69

Similarly, ChemBio Diagnostic Systems Inc. (Medford, NY) introduced their Dual Path Platform Technology, which enabled sequential and orthogonal delivery of sample and detection reagents to the test line in a T-shaped LFA format. They reported an 8-fold sensitivity improvement in a tuberculosis LFA compared to a standard strip format.70 Cho et al. reported an orthogonal delivery scheme to implement an enzymatic amplification step that yielded 30-fold improvement over a conventional Hepatitis B surface antigen LFA.71 However, this scheme was not automated and required an additional user step, which is a common limitation of signal amplification schemes, as discussed below.

Alternate formats exist for other test types as well, such as serological LFAs. For example, Sotnikov et al. performed a comparative theoretical and experimental study for two alternate LFA formats applied to the serodiagnosis of human pulmonary tuberculosis.72 Detection of low-affinity antibodies in human sera samples was 6-fold more sensitive for the alternate format predicted by their theoretical model.

Capture reagents. In addition to non-standard geometries and arrangements of materials, the non-standard arrangement of immobilized capture reagents has been explored as a strategy for improving sensitivity. In a standard LFA, a test line is formed by capture antibodies that have been striped onto the nitrocellulose and allowed to physically adsorb to the membrane. Therefore, researchers have sought to improve immunoassay performance through the controlled orientation of immobilized antibodies; methods of oriented immobilization and characterization were recently reviewed.73 In a different approach, Miller et al. implemented orientation and high-density absorption in a cellulose immobilization system, in which the capture reagent was a highly expressed, recombinant fusion protein containing an analyte-binding domain and a cellulose-binding domain.74 Although this paper-based system was not compared to a standard LFA, the authors reported nearly 100% capture of their target analyte.

Oh et al. refactored a competitive cortisol LFA such that capture reagents replaced the test and control lines to form “deletion” and “detection” zones, respectively. The deletion zone comprised cortisol–BSA conjugates, which captured detection reagents (anti-cortisol antibodies with gold labels) not bound to cortisol analyte. The detection zone comprised anti-mouse IgG antibodies, which captured detection reagents that flowed through the deletion zone due to cortisol analyte binding to the anti-cortisol antibodies. By tuning the concentration of cortisol–BSA conjugates in the deletion zone, and applying a ratiometric imaging method, the authors reported a 30-fold improvement over a conventional cortisol LFA.75

Limitations. The reaction/transport category is home to efforts that use porous materials to automate new reaction schemes with unconventional transport dynamics. As with the innovations in the reaction category, larger gains in sensitivity are likely to be found within the reaction/signal category, although the approaches here are fundamental to mitigating the workflow-related drawbacks of approaches within the reaction/signal category. Thus, we look to this category to provide automation tools within the lateral flow format that will support transformative innovations that will fit in the reaction/transport/signal category. The commercialization of an alternate reagent delivery system (i.e. the Dual Path Platform Technology by ChemBio Diagnostic Systems Inc.) is an encouraging sign that alternative reaction/transport formats are commercially viable.

5.2 Transport/signal

The integration of reagent transport and signal generation in a standard LFA provides an automated mechanism for accumulating signal at a test line. However, the sensitivity of LFAs can be degraded by undesirable noise arising from the incomplete transport (i.e. non-specific binding or adsorption) of detection reagents. Detection reagents immobilized at and around the test line, but not by analyte, represent increased signal noise, which decreases the range of the actual signal. The reduction of sensitivity by assay interferents in human clinical sample matrices, such as blood and urine, due to the non-specific immobilization of detection reagents, has been reviewed extensively.76–79

Blocking. An almost universal strategy for reducing the impact of non-specific binding or adsorption on signal is to dry passive blocking reagents into an LFA. Examples of commonly used passive blocking reagents are bovine serum.
albumin (BSA), casein, antibodies, and/or serum, which are preferred for their low cost and high availability.

More recently, active blocking reagents were developed to further reduce noise, and thus improve sensitivity in LFAs. Commercial examples of active blocking reagents include MAB33 IgG1/IgG1 Poly from Roche Custom Biotech (Indianapolis, IN), TRU Block from Meridian Life Science, Inc (Memphis, TN), HAMA Blocking Reagent from Fitzgerald Industries International (Acton, MA), and the recently launched Morffi from BBI Solutions (Cardiff, UK).

Todd et al. tested HeteroBlock (Omega Biologicals, Bozeman, MT), heterophilic blocking reagent (HBR, Scantibodies Laboratory, Santee, CA), and immunoglobulin-inhibiting reagent (IIR, BioIVT, Baltimore, MD) for depletion of serum rheumatoid factor, using multiplex immunoassay platforms that detect cytokine and chemokine analytes. These tests showed a significant improvement in signal when the active blocking reagents were used. Generally, active blocking reagents have formulations recommended for LFAs or rapid tests, but the appropriate blocking reagent and blocking conditions will depend on the clinical sample type, sample processing methods, target analyte, and other LFA characteristics.

Hybrid materials. The use of nitrocellulose as the analytical membrane in LFAs represents a set of performance trade-offs between transport and signal. The advantage of instrument-free transport provided by capillary action in a porous membrane is offset by the reduction in signal-to-noise ratio created by the non-specific adsorption of signal-generating reagents and the opacity and roughness of the membrane. Fuchiwaki et al. recently introduced an alternative transport system that used a plastic microfluidic channel in place of the analytical membrane and a plastic inlet to produce pressure-driven fluidic transport. They tested the performance of their system against a conventional C-reactive protein LFA and reported 1000-fold sensitivity improvement in their system over the LFA.

Limitations. The transport/signal category contains complementary approaches to the reduction of optical noise in LFAs. The first approach is to minimize the non-specific binding of signal-generating reagents in the analytical material. The second approach is to minimize the background signal of the analytical material itself. The use of hybrid paper- and plastic-microfluidic systems appears to combine the best of both approaches. A plastic microfluidic channel that is low adsorbing and optically clear may be more broadly useful than a blocking solution applied to nitrocellulose. Generally, the limitations of transport/signal approaches are related to the degree to which they require extensive optimization against assay reagents, buffers, and samples. The degree of biochemical complexity found in samples like blood, urine, plant material, stool, etc. is high and drives much of the current need for optimization in LFA development. Thus, an innovation in the transport/signal category will likely need to be completely independent of assay conditions and provide an optimization-free solution—both difficult requirements to meet in a fully integrated LFA format—to be truly impactful.

5.3 Reaction/signal

The integration of assay reactions and signal generation in a standard LFA is achieved in a detection reagent comprised of analyte-binding antibodies conjugated to a colorimetric label. However, a common strategy for improving sensitivity in LFAs is signal amplification, whereby alternative detection reagents support additional reactions that substantively enhance signal. Various approaches to implementing signal amplification in LFAs include the addition of enzymatic, chemical, aggregative, or proximity-dependent reactions.

Enzymatic. Enzymatic signal amplification reactions, which use enzyme labels with convertible (e.g. colourless to coloured, or non-fluorescent to fluorescent) substrates, are well known to improve sensitivity in ELISA, Western blot, and other assays. Horseradish peroxidase (HRP) is a popular enzymatic label used regularly in ELISAs, and was reported in the LFA format as early as 2006 to improve sensitivity 10-fold for both carbaryl and endosulfan in a multianalyte LFA. Many variations on this enzymatic signal amplification reaction are reported in the literature. For example, Parolo et al. modified analyte-specific antibodies with HRP, then functionalyzed gold nanoparticle labels with the modified antibodies to create an amplifying detection reagent. LFAs with these detection reagents were run and visualized using the gold labels, then one of three different colorimetric HRP substrates were run across the test line and visualized again. The authors achieved up to one order of magnitude sensitivity improvement after signal amplification. That basic enzymatic signal amplification approach can be generalized to most detection reagents, for example, a gold label functionalized with nucleic acid probes and HRP, which yielded to a 100-fold sensitivity improvement over the non-HRP detection reagent in a nucleic acid LFA.

Strategies for improving reaction and signal performance separately can be applied to improving integrated reaction/signal performance as well. For example, improving the enzyme amplification reaction by forming the detection reagent from a biotin–antibody to streptavidin–poly-HRP reaction at the test line allowed a PyHRP2 LFA to show equivalent sensitivity to a commercial ELISA due to the fast, strong, and multivalent biotin–streptavidin interaction. The use of multivalent poly-HRP also contributed to the signal amplification. Likewise, the use of a viral nanoparticle as a multivalent detection reagent scaffold, to which multiple HRPs and multiple analyte-specific antibodies were conjugated, improved sensitivity 100-fold in a norovirus GL1 (Norwalk) LFA and nearly 1000-fold in a bacteriophage MS2 LFA.

While HRP is a popular choice for enzymatic signal amplification reactions, alkaline phosphatase (ALP) was used in an LFA for potato virus X to improve sensitivity 27-fold over a conventional LFA. The substrate nitro blue tetrazolium (5-bromo-4-chloro-3-indoly phosphorylate) was added directly to
the test strip over the test and control lines, where it converted to the insoluble, dark-violet precipitate diformazan. Gao et al. used platinum (coated less than ten atomic layers thick around gold nanoparticle labels) as a peroxidase-like catalyst to enhance sensitivity by 2 orders of magnitude in a human prostate-specific antigen LFA. Similarly, Loynachan et al. used porous platinum core–shell nanocatalyst labels and CN/DAB (4-chloro-1-naphthol/3,3′-diaminobenzidine, tetrahydrochloride) substrate in the presence of hydrogen peroxide to obtain a 20-fold improvement over commercial HIV p24 LFAs when run with human sera in the dipstick format.

While many of these recent signal amplification approaches have used smaller substrate volumes for minimal disruption of the LFA format and workflow, we are unaware of embodiments with dry reagent storage reported in the literature. However, Huang et al. constructed a device that provided long-term wet storage of reagents and automated delivery of those reagents to an influenza A/B LFA after sample introduction. Nonetheless, the additional complexity of signal amplification reactions is a barrier to the widespread adoption of these methods to LFA development.

**Chemical.** Chemical signal amplification by silver enhancement is a well-established method to improve sensitivity in protein gel electrophoresis that has also been used in LFAs. For example, the application of silver stain to the whole strip, a method that enhanced sensitivity 1000-fold in a botulism LFA. There have been attempts to simply the workflow of silver staining, and incorporate this step into the standard LFA workflow. Rodriguez et al. tested immobilization of the silver solution within the LFA, but found sensitivity was worse than immersion of the LFA into the silver solution. Lathwal and Sikes added of a small volume of commercially available silver solution, freshly mixed, but other signal amplification methods described in the same paper had higher assay sensitivity. Panferov et al. also a small volume of silver solution and obtained a 10-fold sensitivity improvement, compared to no stain, in a *Ralstonia solanacearum* LFA. Anfossi et al. were able to increase sensitivity 10-fold in an ochratoxin A LFA with a competitive assay and silver stain, which enabled detection of the toxin in wine and grape musts at levels commensurate with strict European regulatory limits.

All these examples, which clearly demonstrated sensitivity improvements from the use of silver stain on LFAs, nonetheless required freshly made silver solutions to be pipetted onto an LFA at specific times in specific volumes. Fridley et al. dried a commercial gold enhancement (salt-based) solution onto a malaria PfHRP2 LFA, and demonstrated a 3-fold improvement in sensitivity, but required a device that modified transport to achieve this modest result. As with the enzymatic signal amplification approach, reaction/signal modifications can trigger the need to deploy complementary transport modifications as well. We will return to this topic below, where we discuss interconnection of reaction, transport, and signal in LFA systems.

A reported alternative chemical approach is polymerization-based signal amplification (PBA), where the detection reagent contains a photoinitiator that triggers hydrogel polymerization in the presence of light and acrylate monomers; a pH indicator converts the presence of the hydrogel into a colorimetric signal. This signal amplification reaction was introduced as part of a paper-based PfHRP2 assay with a reported LOD of 0.56 μg mL−1. The primary advantage of the PfHRP2 paper-based PBA assay was speed, with a time-to-result of 2–3 minutes, as PfHRP2 LFAs with enzymatic amplification have better sensitivities but with a typical time-to-result of 20–30 minutes.

**Aggregative.** The aggregation of signal-generating labels at the test line is usually entirely due to the base reaction between immobilized capture reagent, analyte, and labelled detection reagent. However, assay modifications that add further aggregation reactions can amplify signal.

For example, a two-gold-label, multi-level aggregation system resulted in a 100-fold sensitivity improvement in a tropinin LFA over the standard, one-gold-label system. In the two-gold-label system, a smaller (10 nm) gold label was functionalized with anti-analyte antibodies and blocked with BSA, then dried downstream of a larger (40 nm) gold label, which was functionalized with anti-BSA antibodies. Upon application to the LFA of a sample containing analyte, the smaller gold labels were pulled down to the test line first, by the analyte. The larger gold labels were then pulled down to the test line second, by the BSA on the smaller gold labels. Two variations on this method were described by Wiriyachaiporn et al., in that the larger gold labels were pulled down with a biotin–anti-biotin antibody interaction and two analytes were detected on the same test line. Capture of a second analyte increased the overall density of labels at the test line. The authors reported an 8-fold improvement in sensitivity over a conventional, single-analyte LFA. Taranova et al. used three labels (functionalized with specific antibody, biotin, and streptavidin, respectively) in a similar format to increase the sensitivity 30-fold in a procalcitonin LFA. Razo et al. used two different label types, magnetic nanoparticles conjugated to biotinylated antibodies and streptavidin-coated gold nanoparticles, and a magnetic concentration step to obtain 32-fold sensitivity improvement in an LFA for potato virus X over a conventional LFA without the magnetic concentration or the gold-label-mediated aggregation.

**Proximity-dependent.** Proximity-dependent reactions effectively amplify signal by creating a signal only from the combination of detection and capture reagents. In these reactions, no signal arises from either reagent separately, and signal is generated only when their separate binding to analyte brings them into proximity. The advantage of this approach is that there is little to no signal generated by nonspecific binding, which enhances the range of the signal, which hopefully improves sensitivity.

A classic form of proximity-dependent reaction is fluorescence (or Förster) resonance energy transfer (FRET), in which energy from a fluorescently excited donor molecule is...
transferred non-radiatively to an acceptor molecule in close vicinity.\textsuperscript{101} This energy transfer significantly reduces the fluorescence signal from the donor molecule, and can increase the fluorescence signal from the acceptor molecule unless it is a fluorescence quencher. Either signal can be monitored with a fluorescence reader.

Recently, Wang et al. reported the use of gold nanoparticle labels to quench fluorescein-labelled capture antibodies in a carcinoembryonic antigen (CEA) LFA.\textsuperscript{102} The sensitivity increased 100-fold from the standard, non-FRET LFA with gold labels only. However, as with any non-colorimetric signal, instrumentation was required to measure the fluorescent signal. Additionally, signal calibrations showed that while the accuracy of the LFA was quite good for real patient samples, the coefficients of variation were high (in some cases greater than 20%).

More complicated reaction mechanisms such as proximity ligation and proximity extension have been recently reviewed.\textsuperscript{103} In these assays, oligonucleotides are conjugated to both capture and detection reagents, such that hybridization between oligos or to a third oligo can occur when capture and detection reagents are both bound to analyte. Proximity-dependent hybridization is followed by ligation or extension, then nucleic acid amplification. This reaction mechanism is sufficiently complicated to challenge the typical LFA workflow, but a variant of the oligonucleotide ligation assay (OLA) was reported with sensitivity equivalent to a plate-based OLA, but also with 3-fold reduction in assay time, in an HIV LFA for the detection of viral point mutations.\textsuperscript{104}

Limitations. The reaction/signal category is mostly composed of approaches that introduce more complicated reaction schemes to an LFA than the conventional sandwich or competitive immunoassay. These schemes amplify a signal target molecule or the signal itself beyond a purely proportional measure of the analyte. The main limitation of these approaches is that more complication reaction schemes are not easily accommodated within the conventional workflow. Additional user steps compromise the overall utility of the LFA, especially in point-of-care applications. Therefore, we believe that adoption of innovations in the reaction/signal category into current practice require related innovations in automated transport systems, such that the most practicable form of a reaction/signal innovation should arise from the reaction/signal/transport category.

5.4 Reaction/transport/signal

The integration of chemical, physical, and optical processes within an LFA produces the major advantages of the technology, which revolve around simple, rapid, and robust operation. However, as we have explored here, this integrated functionality results in a highly complex set of trade-offs that affect overall system performance. As previously mentioned, the approaches we discuss in this review have improved sensitivity relative to baseline performance, but in the process may have increased the number or complexity of user steps, increased overall test time or cost, decreased test robustness, or introduced expensive instrumentation.

To put it another way, we quickly see the distinctions blur between the reaction/transport/signal categories for the more impactful innovations, but also that any modification within one or two of these categories will trigger complex trade-offs across all categories. Even in conventional LFA development, navigating trade-offs in system performance typically requires a long process of manual optimization over many experimental parameters, including those related to reagents and reaction conditions, materials, and test signals.\textsuperscript{105,106}

For example, in conventional LFA development, reagent optimization requires (at least) screening for high-affinity antibodies, selecting colorimetric labels, assessing the efficacy of attachment chemistry, cross-testing the capture and detection reagents, and evaluating the impact of blocking, spotting, drying, and/or rehydrating reagents in porous materials. Reaction condition optimization can include evaluating with sample matrices, and various buffers (for storage, reaction, wash/rinse, blocking, and/or running), surfactants, and other additives; testing the impact of various temperature, salt, and pH conditions; and testing porous materials for specific and non-specific binding. Material optimization typically involves choosing porous material types, width, thicknesses, lengths and overlaps; adding a backing strip and optionally a cover tape; and adjusting the mechanical design of a cassette. During these and potentially other development steps, validation (as far as possible) of optimal conditions at each step should be performed with statistically relevant measurements of signal-to-noise ratios over a relevant range of analyte concentrations, in a realistic sample matrix when possible, which is a huge amount of work.

Automated optimization. One approach to the tackling the laborious and time-consuming process of empirically screening through and optimizing across the copious development options is automation. For example, our FlowDx group at Intellectual Ventures Laboratory is developing an platform to automate a relatively rapid, high-throughput, design-of-experiments (DOE) approach to LFA optimization.\textsuperscript{107} This platform uses a robotic liquid handler to automatically populate a large batch of LFA strips with various assay conditions at the test and control lines, in the conjugate pad, and on the sample pad. It then automatically runs the tests with the appropriate buffers and samples and images the results. This automated system will considerably improve testing throughput. In combination with the statistical power of the DOE approach, response surfaces for optimal LFA behaviour can be created and used to recommend directions for LFA optimization.

Computational modelling. Another approach to the systematic improvement of LFAs, including sensitivity enhancement, is the application of computational models to the LFA development process, which is an area recently reviewed in detail.\textsuperscript{108} The diagram in Fig. 2 illustrates how LFAs are complex systems composed of many different biological and structural features. Mathematical models should consider
flow properties, membrane porosity, surface interactions, as well as receptor-ligand kinetic interactions, and the presence of non-specific binding components.

One of the earliest LFA modelling approaches predicted the non-linearity of signal response at high target analyte concentration in an LFA sandwich format, and defined reactions and equations to allow for computational—rather than empirical—optimization.\textsuperscript{105} Recent advances in LFA modelling include work by Ragavendran et al. to optimize placement of the test line on the LFA strip and sample volume,\textsuperscript{110} and by Sotnikov et al. to optimize serological LFAs in which the target is an antibody, rather than an antigen.\textsuperscript{111} For the latter, the presence of excess target antibody in the blood sample adds an additional layer of complexity to the LFA assay; the authors recommend high concentration of reagents and serum dilutions in the range of 10–100-fold to improve sensitivity.

Zhan et al. recently modelled and experimentally validated the effect of gold nanoparticle sizes on detection of an LFA signal by a thermal contrast reader.\textsuperscript{112} The model and empirical data showed that larger particles increased antibody load due to their higher surface area, and gave stronger thermal responses. The use of 100 nm gold labels in their LFA led to a 256-fold increase in assay sensitivity compared to 30 nm gold labels.

Recently, we have developed and validated a mechanistic model in the FlowDx group at Intellectual Ventures Laboratory to broadly improve LFA sensitivity.\textsuperscript{108} The model accepts input parameters that include kinetic binding constants for reactions, porosity of transport materials, and viscosity of fluids, which allows for \textit{in silico} optimization of signal generation and/or reagent use. Furthermore, the model can identify certain types of non-specific binding, which allows for at least partial \textit{in silico} optimization of noise reduction. To date, we have validated the model for a malaria LFA under development.

\textbf{Limitations}. In a sense, the reaction/transport/signal category encompasses all the innovations we have included in this review because—although we have used subcategories to emphasize the focus of individual innovations—the interconnectedness of LFAs means that any modification of an LFA can have a system-level impact on functionality and performance. The result, with respect to innovations that lead to improvements in LFA sensitivity, is that modifications must be exhaustively tested and optimized (over antigens, samples, reagents, materials, etc.) to prove system functionality and performance, or must be paired with complementary modifications to preserve the same. Therefore, what we decided to include in the reaction/transport/signal category for the purposes of this review are systematic approaches to LFA development that might eventually remove these burdens on integrating innovative modifications into LFAs. The qualification on the two approaches that we included is that neither is currently fully realized. Full realization of either approach is limited mainly by the level of investment required to address all aspects of an integrated LFA system. For example, the automation of LFA optimization that we described was enabled through modification of an expensive, general-purpose, liquid-handling robot, which is likely not possible to modify in a way that would enable automated manufacture and construction of LFA test strips from raw materials. Unless automated manufacturing equipment is incorporated into the system, a significant portion of this modified LFA development process will remain laborious and time-consuming. Likewise, predictive design tools that address portions of the LFA development process are beginning to emerge that are based on state-of-the-art computational models of LFA systems. However, to address the entire development process, and any innovative modifications thereof, more sophisticated multimodal and multi-scale models are required.

\section*{Conclusions}

We have used a categorical framework to identify state-of-the-art innovations in LFA sensitivity enhancement, group them according to shared strategies, and critically assess their limitations from a system-level perspective. We have also argued that a system-level perspective is necessary to overcome limitations blocking many of these innovations from transforming the current practice of LFA development. In anticipation of this outcome, we look forward to a new generation of LFAs that have been systematically improved to remain specific, stable, rapid, inexpensive, and user-friendly; and to acquire sensitivities that rival expensive, instrumented, laboratory-based assays.

\section*{Conflicts of interest}

There are no conflicts to declare.

\section*{Acknowledgements}

Our thanks to our colleagues in the FlowDx group at Intellectual Ventures Laboratory, as well as to our partners at Diagnostic Consulting Network, Carlsbad, CA, USA. Funding provided by The Global Good Fund I, LCC (www.globalgood.com).

\section*{Notes and references}

\begin{enumerate}
\end{enumerate}
Critical review

10 S. Y. Doerflinger, J. Tabatabai, P. Schnitzler and C. Farah, mSphere, 2016, 1, 1–6.