Perspective review of what is needed for molecular-specific fluorescence-guided surgery

Brian W. Pogue
Eben L. Rosenthal
Samuel Achilefu
Gooitzen M. van Dam
Molecular image-guided surgery has emerged as a very active research field with transformative potential to allow surgeons to see the molecular phenotype or even genotype of diseased and normal tissues during surgery.1–3 Much of this field’s clinical foundational roots have been developed based upon vascular perfusion imaging with indocyanine green,4–6 with significant growing interest in the use of other Food and Drug Administration (USFDA)-approved agents, such as sodium fluorescein, methylene blue, and isosulfan blue, which can have different transport and localization kinetics.7–12 Also related are the use of endogenous signals such as autofluorescence attributed to NADH/FAD concentrations, as well as absorbers such as hemoglobin and lipids,13 and scattering agents, such as endogenous4,14 or surface-enhanced Raman scattering.15 There have also been decades of studies and approvals in photodynamic agents for diagnostic use in surgery, using porphyrins, phthalocyanines, and chlorins as uptake markers of tumors.16–21 A major milestone in this field just occurred, which was the USFDA approval of aminolevulinic acid-induced protoporphyrin IX to guide neurosurgery.22 Although already used in Europe, this action represents the first biosynthetically activated molecular probe approved for human use to guide surgery in the United States (US), named Gleolan (NX Dev Corp., recently acquired by photonamic GmbH & Co. KG). While 5-aminolevulinic acid hydrochloride (ALA) has been used in human research and approved for human use for diagnostics for decades, this recent approval comes from the result of a Phase III trial, is industry supported, and is independent of the imaging system used. This combination of features marks a very important turning point in the field. However, as a field of research in the US, surgical guidance is highly delocalized and broken down into subspecialties, without significant intercommunication, even though there is a high level of activity in each discipline. So, while many research groups are poised to achieve major discoveries and seminal studies in surgical trials, yet there is also some sense that there are major barriers to seeing this happen. This review focuses on the key aspects that require care in the clinical translation pipeline, or aspects that may be limiting the field today, and some major threats that could potentially stop the advancement of the field if not solved.

Several key factors influencing the field are shown in Table 1, broken down by analysis of strengths, weaknesses,
opportunities, and threats (SWOT paradigm). Many of the strengths of the field have historically been described as needs or opportunities, but at this time, there have been major improvements that can now be counted as strengths of the field. Each of the four areas is discussed in detail in the following paragraphs.

Arguably, one of the most important strengths of the field today is the commercial interest in providing imaging systems, leading to prototypes with strong ergonomic features for clinical use. The lack of choice and features in these systems has been a problem in the past, but today there are many systems that have been steadily improving. This success has largely been due to resurgence in the use of indocyanine green (ICG) for tissue perfusion assessment, largely driven by Novadaq Tech. Inc. (now acquired by Stryker Corp.) pushing the field forward to evolve the commercial market for this type of direct visual guidance. This has enticed production of multiple systems from manufacturers with more differentiating features to the original systems and adoption into many different surgical specialty systems. There is clearly concern about variation in system capabilities and the potential confusion of systems, and this is a topic of ongoing study in the technical performance, calibration, references standards, professional consensus, and regulatory guidance. This ongoing evolution and discussion at a high level of technological activity should be viewed as a strength of the field. However, much of the activity is based on anticipated success in the development of companion diagnostic agents that are highly sensitive, nontoxic, and can fit into the workflow of surgery. Similar strengths are in the early entry into exploratory uses of ICG in Phase 0/1 trials in a number of centers, as well as first in human studies of new molecules (shown in Table 2). The key major opportunity for this field lies in the widespread adoption of molecular pathology and proteomic/genomic phenotyping of diseases in all major care centers. This means that, currently, surgeons have more knowledge than ever before about what the phenotype of the tissue is that they will be operating on, and this field is growing at an explosive rate. It is very likely that within the next decade that all tumors will have full profiling from mRNA, DNA, or proteomic assay. However, until this phenotypic profiling is truly used in surgical guidance, it remains an undeveloped opportunity. Still the idea of using the most specific proteins as carriers of molecular probes to deliver them to target tissue is extremely attractive. Particularly, the cancer therapeutics world has solid progressive growth-targeted inhibitors or proteins that could utilize a fluorescent companion diagnostic agent. The concept of adopting therapeutic antibodies as the fluorophore-carrying moiety, thereby using therapeutic biologicals as imaging agents, has already been demonstrated in ongoing surgical trials. Further growth along this direction is very likely going to happen and will capitalize on this approach. The opposite opportunity is also possible, that molecular imaging agents delivered prior to surgery can provide unique information about the tumor susceptibility to certain targeted drugs, susceptibility to systemically delivered agents, and appropriate therapeutic agent dosing. As the agent would be delivered prior to surgical removal, the tissue would be available for molecular, genomic, and histological examination.

The other major opportunity has been the creative growth in molecular probes for surgical guidance. A listing of those registered with clinicaltrials.org for patient recruitment is shown below in Table 2, illustrating the wide diversity of probes being tested, with a range of more and more specific binding
<table>
<thead>
<tr>
<th>Type of molecular probe</th>
<th>Molecule/probe</th>
<th>Commercial name</th>
<th>Clinical site or purpose (registered at <a href="http://www.clinicaltrials.gov">www.clinicaltrials.gov</a>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascular perfusion/flow</td>
<td>Indocyanine green</td>
<td>ICG, AIM ICG</td>
<td>Many organ/tissue sites (&gt;400 trials)</td>
</tr>
<tr>
<td>Lymphatic flow/sentinel nodes tissue retention</td>
<td>Indocyanine green</td>
<td>ICG</td>
<td>Breast, parathyroid, tumors (&gt;40)</td>
</tr>
<tr>
<td></td>
<td>Methylene blue</td>
<td>MB-102</td>
<td></td>
</tr>
<tr>
<td>Autofluorescence</td>
<td>NADH/FAD</td>
<td>n/a</td>
<td>Many organ/tissue sites (&gt;190 trials)</td>
</tr>
<tr>
<td>DNA intercalation</td>
<td>Prolavine</td>
<td>n/a</td>
<td>Squamous cell neoplasia, Barrett’s esophagus, colon polyps, dysplasia, anal dysplasia, head and neck cancer, cervix cancer, uterine cancer, oral disorders, gastric cancer (17 trials)</td>
</tr>
<tr>
<td>Molecular vibrations</td>
<td>Raman scattering</td>
<td>n/a</td>
<td>Liver, macula, foot ulcers, glucose (&gt;100 trials)</td>
</tr>
<tr>
<td>Metabolism—enzyme or synthetic activity</td>
<td>ALA</td>
<td>Gliolan</td>
<td>Glioma, bladder (13 trials)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Levulan</td>
<td>Skin precancers and cancers (&gt;200 mostly PDT trials)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPC-07</td>
<td>Glioma (one trial)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolism—carbohydrates and proteins</td>
<td>Hexaminolevulinate</td>
<td>Cysview</td>
<td>Bladder, cervix, colorectal cancer (four trials)</td>
</tr>
<tr>
<td></td>
<td>Methyl aminolevulinate</td>
<td>Metvixia</td>
<td>Skin AKs, cancers, Bowen’s disease, acne (68 mostly PDT trials)</td>
</tr>
<tr>
<td></td>
<td>Cathepsin activatable</td>
<td>LUM015</td>
<td>Sarcoma, colorectal, pancreatic esophageal, breast, prostate cancers (five trials)</td>
</tr>
<tr>
<td></td>
<td>Protease activatable</td>
<td>AVB-620</td>
<td>Breast cancer (two trials)</td>
</tr>
<tr>
<td></td>
<td>Fluorescent lectin</td>
<td>n/a</td>
<td>Colorectal cancer, neoplasms, polyps (one trial)</td>
</tr>
<tr>
<td></td>
<td>HSP90 inhibitor</td>
<td>HS-196</td>
<td>Solid tumors (one trial)</td>
</tr>
<tr>
<td></td>
<td>Chlorotoxin blocking</td>
<td>BLZ-100</td>
<td>Glioma, breast, CNS, skin, sarcoma (five trials)</td>
</tr>
<tr>
<td>Immunology—receptor and cell surface protein targeting</td>
<td>7-aa peptide—IRDye800CW</td>
<td>KSP-910638G heptapeptide</td>
<td>Gastrointestinal malignancies</td>
</tr>
<tr>
<td></td>
<td>c-Met targeting peptide</td>
<td>EMI-137</td>
<td>Colon cancer, esophageal cancer and high grade dysplasia, papillary thyroid cancer, lung cancer (four trials)</td>
</tr>
<tr>
<td></td>
<td>Folate receptor targeting</td>
<td>OTL38</td>
<td>Renal cell, lung, ovarian, pituitary, pleural cancers (nine trials)</td>
</tr>
<tr>
<td></td>
<td>Tumor-specific integrin receptor binding</td>
<td>LS301</td>
<td>Breast cancer (one trial)</td>
</tr>
<tr>
<td></td>
<td>Anti-EGFR binding peptide</td>
<td>QRH-882260</td>
<td>Colon cancer, cholangiocarcinoma (three trials)</td>
</tr>
<tr>
<td></td>
<td>Anti-EGFR affibody</td>
<td>ABY-029</td>
<td>Glioma, sarcoma, head, and neck (three trials)</td>
</tr>
<tr>
<td></td>
<td>GRPR receptor binding peptide</td>
<td>^66GA-BBN-IRDye800CW</td>
<td>Glioblastoma (two trials)</td>
</tr>
<tr>
<td></td>
<td>VEGF antibody</td>
<td>Bevacizumab-IRDye800CW</td>
<td>Esophageal, breast cancer, adenomatous polyposis (nine trials)</td>
</tr>
<tr>
<td></td>
<td>EGFR antibody</td>
<td>Cetuximab-IRDye800CW</td>
<td>Pancreatic cancer, brain neoplasms, glioma, head and neck squamous cell carcinoma, head and neck cancer (four trials)</td>
</tr>
<tr>
<td></td>
<td>EGFR antibody</td>
<td>Panitumumab-IRDye800</td>
<td>Pancreatic cancer, brain neoplasms, glioma, head and neck squamous cell carcinoma, head and neck cancer (four trials)</td>
</tr>
<tr>
<td></td>
<td>Carbonic anhydrase IX antibody</td>
<td>^111In-DOTA-Girentuximab-IRDye800CW</td>
<td>Renal cell carcinoma (one trial)</td>
</tr>
</tbody>
</table>
characteristics, ranging from endogenous molecules in the beginning to exogenous probes that specially bind to cell surface carbohydrates, free proteins, specific enzymes, cellular channels, or cell surface receptors. The specificity of each is known a priori, and the efficacy in vivo is being tested in these early phase trials. At this point, most of the more specific agents have not advanced beyond Phase 1 trial, but there is potential. However, some of the factors listed next will determine success or failure of these trials.

The weaknesses of the field are the issues that will limit the successes if not solved. One of the most obvious and yet least discussed aspects of the field is that molecular-specific dyes have a localization that is often microscopically heterogeneous, and so this is not easily appreciated with surgical imaging in vivo. Although bulk tissue removal can still aid the surgeon as guided by the fluorescence signal, there are diseases, such as brain or ovarian cancer, where microscopic pathology removal is more critical and macroscopic views of the fluorescence will be insufficient. Pathology analysis of immunohistochemistry slides routinely shows high morphological heterogeneity in tumor nodules and ducts, indicating that specific molecular probes to these targets would likely also be microscopically heterogeneous. A pathologist is trained to view this heterogeneity and make sense of it, knowing that staining patterns are usually complex, and it is this complexity that can be one of the keys to diagnosis. Yet a surgeon must perform resection at the macroscopic or perhaps mesoscopic resolution. The macroscopic view of a heterogeneous uptake blurs out the uptake and observed contrast, and so in vivo levels of tumor to background are routinely seen near 1.1x to 5x.35,36 Whereas if the view was at the cellular or near cellular level, contrast appears at much higher levels, nearer 10x to 30x is common, but this cannot be resolved when imaging at the macroscopic scale (see examples from three different molecular tracers in Fig. 1). The difficulty lies in the basic fact that most surgical procedures perform random sampling of the tumor at the microscopic level through use of frozen sections, but do not allow for thorough microscopic examination of the tissue, and the accompanying microscopic-level image extraction. The surgeon, in general, is interested in the margin and not the complete volume of the tumor, even with all its heterogeneous components. It is the microscopic margin and the interaction with the microenvironment that presents the leading challenge to the surgeon toward macroscopic imaging. So, the field is left with a situation where the technological imaging tools used will not allow for imaging of the true disease-specific morphology, which might be better used. Perhaps even worse is the fact that as this is seemingly unrealizable, most of the focus of the field is on quantifying the available modest contrast and making decisions based upon this rather poor reporter of the true molecular specificity.40 An interesting side aspect of this though is that this problem is an issue for nuclear medicine molecular imaging as well as fluorescence-guided surgery (FGS), as macroscopic imaging resolution simply does not allow for delineation of the microscopic phenotypic heterogeneity of cancer, whereas in standard-of-care one might challenge if this is a real problem in the surgical theatre.

One solution to imaging of the cancer heterogeneity issues, is to have a combination of: (1) a widefield high-resolution imaging system and (2) an automated software to recognize and classify the molecular morphology patterns. This would be similar to what is seen in immunohistochemistry, and automation of this is coming, but for now it remains one of the key weaknesses of the field, where the observed contrast levels seen are distinctly unimpressive. Alternatively, it is possible that fast magnification change imaging could be possible where the low power, low magnification can be quickly swept up to higher power and high resolution. The value of this is that the scoping in and out from macroscopic to microscopic imaging would allow for superior workflow for the surgeon, who desires the ability to see boundaries or fine structures better. Without these types of tools, many companies developing molecular probes may fail to see the high contrast in vivo, ignoring the tumor specificity it delivers in ex vivo analyses. As such, they might see a mismatch between in vivo and in vitro data and stop development of new molecular probes. So, while this problem can seem trivial, without a technological or logistical solution to it, there will remain a lack of clarity in target validation. What remains pivotal is comparing imaging data with the accepted gold standard in daily clinical care, which remains simple H/E histopathology. One practical issue is the comparison of frozen section fluorescence imaging to fixed H&E pathology images, which requires challenging and labor-intensive comparison studies.35,41,42 This is done because many targeted probes do not remain intact or active when fixed, although there are some which appear to have this characteristic. This seemingly trivial issue is a major challenge though to target verification with in vivo–ex vivo assay. Efforts to apply deep learning systems comparing in vivo fluorescence with ex vivo tumor delineation might be a solution in this complex equation. This is an important part of clinical trial testing, which without an appreciation for it could lead to a slowing of enthusiasm because of a perceived low contrast, stymying research and development in this field.

Another major weakness of the field, which pervades all of molecular cancer research, is the over-reliance of preclinical testing in tumor cell lines that are monolithically positive for the molecular target of interest.35,44 For example, if a single receptor directed probe is being tested, then it is tested on a tumor line that has exceptionally high expression of that receptor, because it needs to be shown that it has specificity. However, when deployed in humans, the tumor heterogeneity in terms of any protein expression can range by orders of magnitude. This preclinical approach is done for practical and economic reasons, and even though it is a widely recognized problem in cancer research, and publishers and funding agencies are trying to mitigate it, there are few obvious practical solutions.45 The cost of preclinical work is high, and the complexity of trying to test a molecular probe in hundreds of tumor models in vivo is implausible, although some methods are being proposed now.46 Yet, as soon as the same molecular probe is used in human trials, the range of tumors imaged can be enormously more variable than the singular xenograft models that were tested on. Additionally, the size and growth of the human tumor can alter the value of targeting, as it is well known that as tumors grow, the phenotypic characteristics can vary throughout the tumor, with variations in gene expression, protein expression, and mutated protein expression.

Given this, the ability to know the expression level and the heterogeneity of expression may be the key to allowing success in use of the molecular probe.43 A concern is the use of targeted fluorescence imaging in a clinical trial, often consisting of a heterogeneous patient population in terms of age, sex, tumor size, tumor heterogeneity, neoadjuvant treatment, comorbidities, etc., which might have results that are confounded by uncertain
or inconclusive data about the observed target contrast, without better interpretation of both the expression and microscopic heterogeneity of this expression. Mitigating this problem could come at both ends of the pipeline, in terms of using more tumor lines in three-dimensional (3-D) culture or premade arrays, and then also better appreciating the in vivo pathology of each individual patient being imaged. Measurements of molecular probes though are limited to a static IHC expression,
and this expression could vary over time and with sample handling. Target expression that is expressed at a low level but internalizes may produce a much better signal. Furthermore, establishment of the need to identify the marker on every patient will make the process more expensive and invasive (especially, in sites like brain tumors where biopsies are rarely performed). In the end, the only current plausible solution is that the probe validation needs to be performed postsurgery by assessment of the histology/fluorescence.

Another weakness of the field is the regulatory processes that limit development and deployment of imaging probes to one at a time, with good manufacturing practices (GMP) production requirements, release testing, toxicity testing, and ongoing stability analysis for each molecule. It is possible that optimal tissue selectivity may require the type of combination targeted which is common place in clinical oncology today. This could take advantage of multiple epitopes, multiple receptors, and normal and mutated proteins, protein–carbohydrate combinations, multiple enzymes, or combinations of metabolism and immune profiles if tissue. However, the commercial barriers to imaging agents are significant, and even getting a single agent into approved/cleared use today seems nearly insurmountable. Thus, without a different strategy, multiple targeting moieties are unlikely to be commercially plausible. In fact, the codelivery of a therapeutic and diagnostic agent is a conflict of interest as the returns are so lopsided; industry is often unwilling to tolerate the risk of the therapeutic agent.

The threats to FGS are perhaps the most troubling and require some careful thought before they derail progress in the field. Although each threat may not have immediate solutions, it is critical to appreciate them. The largest and most pressing threat is the fact that clinically observed contrast and uptake data are being misinterpreted at times in terms of the specificity of the signal. As previously mentioned in the weakness area, limitations of current macroscopic imaging systems can be a problem that leads to an underestimation of the specificity for the molecular probe to the target. But an even further problem that is more insidious is that macroscopic imaging tools simply cannot provide verified information about the specificity of uptake. It is very common for researchers to interpret the macroscopic contrast as the specificity of probe, and yet, factors such as vascular perfusion and lymphatic impairment can often have more effect upon the uptake than any specificity of binding or localization. This is well known as the enhanced permeability and retention (EPR) effect, and yet it is also commonly ignored when interpreting preclinical data and clinical data. This is avoided often for the simple reason that the macroscopic imaging tool cannot resolve the uptake and that the uptake patterns are mixed in with vascular perfusion patterns. Although additional target validation at the microscopic level is possible, it requires additional effort and is not always recognized by people in the development pipeline as a necessary step in the process of agent development. As a result, parsing out the vascular, localization, and lymphatic effects can be scientifically challenging and expensive to do well (Fig. 2).

It is very common to simply compare two tumor lines, one positive for the target, and one negative for the target as a quantitative assay of specificity of uptake. Yet, while this seems reasonable, these two tumor lines can easily have much larger vascular perfusion and lymphatic flow differences, which will dwarf the effects of the probe binding. Preclinical testing of specificity in this manner, while seemingly logical, based just on the target, adds a high level of randomness to the field, and many published studies have data that are confounded by this issue. Solutions to this problem are complex and require more careful microscopic and pathologic analysis of uptake, comparing matched targeted and untargeted probes, to ensure that the measured bulk uptake is actually bound to the target or activated by the expected target. Solutions such as using a reference tissue to normalize are insufficient, because they do not solve the problem of the specificity of uptake. Modeling studies, detailed immunohistochemistry studies, or comparison with

Fig. 2  Example of microscopic target validation: panitumumab-IRDye800 localizing in head and neck squamous cell carcinoma (yellow line) but not in surrounding stromal and inflammatory tissues. This represents positive target validation but does not rule out off-target effects or failure to target all elements of cancer distributed throughout the tumor mass.
a paired tracer can each potentially mitigate the problem, but are more easily implemented at the preclinical stage than the clinical stage.

Another key threat is the lack of an obvious way to advance production and in vivo toxicity testing in a way that allows the cost to match the benefit. The hardest part of this as the field evolves is having examples of business cases that allow for companies to recoup their expenditures for the development cost. Current GMP (cGMP) production of small organic dyes for human use have been pioneered by a few groups, such as the work of Hyun and colleagues. Conventional contrast agents have been reducing in numbers for most of the last decade, and the development of new CT and MRI contrast agents is comparatively stagnant, and this is even true in nuclear medicine. This is related to a lack of additive potential revenue from their use. This is a challenging problem, although solutions may exist in following a few of the case examples in nuclear medicine, taking advantage of the use of cGMP methods in academic laboratories with initial human testing in microdose trials. Perhaps the most common method of using this research-based approach, especially when there is limited return but high value to the patient, is the development of registries supported by the Centers for Medicare and Medicaid Services. This method was critically successful in molecular imaging, providing for the development of FDG-PET imaging, and seems a viable pathway, as long as the imaging systems used are sufficiently sensitive enough to image nanomolar levels of agents. Other solutions may also emerge as the field develops, such as gaining approvals for mixtures of agents at a time, although this is seemingly harder to imagine as time goes on.

Toxicity and release testing will always be a necessary and expensive part of the process as well, which cannot be avoided. However, there are methods that can limit costs, if, for example, the imaging is done with therapeutic proteins that have already undergone substantial human use, and can be slightly modified without altering their pharmacokinetic behavior. In these cases, minimal testing might be feasible under negotiation with regulatory bodies. Alternatively, microdose studies are also possible, which reduce the required toxicity studies to a single rodent species with a single injection, from the more complex multiple-species-multiple-dose studies. This can be effectively completed for a modest cost and can even be completed on the test batch of an agent instead of the final product, as long as there is consistent good laboratory practice followed between the production stages. Toxicity testing will become a larger aspect of agents as they expand and while this is not a biomedic optics problem per se, there are approaches such as in vivo pharmacokinetic imaging, which might be deployed to reduce animal use and cost. Similarly, in vivo optical imaging to assess activity and target bioavailability will likely be useful to confirm the value of agents. Finally, better understanding of the in vivo kinetics and clearance organ behavior of both the dye and the targeting moiety can alleviate some of the costs of preclinical work, where some are developed without sufficient planning for avoiding organs at risk or appropriate clearance kinetics.

The threat about economic viability for new molecular probes is less of an issue for small molecules than it is for larger biological agents, and it is likely that the two will have very different development paths. It is possible that once the safety of fluorescently labeled therapeutic antibodies has been better established, one could make the case that small labeling changes are relatively minor, and that these agents, if made under certain parameters, should inherently have FDA clearance if the antibodies are already in common human use. This is valid if the fluorescent labeled antibody has the same biological clearance and activity as the native antibody, because with careful labeling there is minimal effect to a 155-kDa antibody from a 0.5-kDa dye added to it. This has been done by several groups in ongoing clinical trials in the US, with Cetuximab and Panitumumab, and in Europe, with bevacizumab and an anti-CEA antibody. The application of this approach to developing agents for multicenter clinical trials has not been achieved yet, nor to FDA cleared products, but it is a promising pathway.

**Summary**

The perceived value of molecular guided surgery is exceptionally high. It comes with the promise of combining molecular pathology information into a nuclear medicine-like molecular guidance, which would improve the precision of surgical procedures. Increasing interest by different surgical specialties has been robust, matching the commercial investment occurring. To keep this field progressing to its highest potential, the issues mentioned in the weaknesses and threats areas need attention and creative input from partners in engineering, medicine, and industry.

**Disclosures**

The authors have no funding, financial relationships, or conflicts of interest to disclose. The authors have financial conflicts of interest to disclose related to the content of this article.

**Acknowledgments**

This work has been supported by NIH Grant Nos. R01 CA167413 and R01 CA190306.

**References**


Biographies for the authors are not available.