PNAS
www.pnas.org

Supplementary information for

Immunomagnetic microscopy of tumor tissues using quantum sensors in diamond

Sanyou Chen, Wanhe Li, Xiaohu Zheng, Pei Yu, Pengfei Wang, Ziting Sun, Yao Xu, Defeng Jiao, Xiangyu Ye, Mingcheng Cai, Mengze Shen, Mengqi Wang, Qi Zhang, Fei Kong, Ya Wang, Jie He, Haiming Wei, Fazhan Shi, and Jiangfeng Du

Fazhan Shi
Email: fzshi@ustc.edu.cn
Jiangfeng Du
Email: djf@ustc.edu.cn

Supplementary information section

1. Simulation of the nitrogen distribution in a 100-nm NV layer diamond

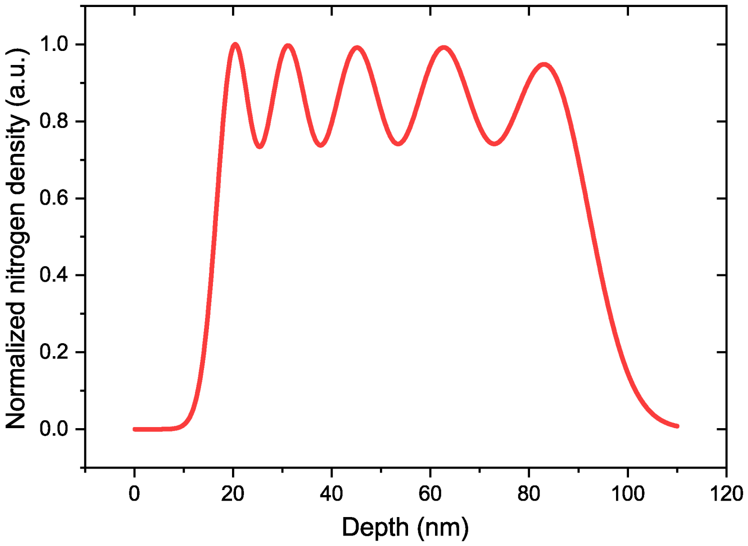

Fig. S1. Simulation of the nitrogen distribution in a 100-nm NV layer diamond. The almost homogenous nitrogen with a depth range between 10 nm and 110 nm below the diamond surface was estimated by Stopping and Range of Ions in Matter (SRIM) Monte Carlo simulation. The 100-nm nitrogen layer can be created by a multi-implantation of $14N^+$ ions. The implantation sequentially goes through sequential five energies with corresponding five different doses (See Methods). After once annealing, we can obtain NV centers with a similar distribution between 10 nm and 110 nm.

2. Magnetic signal simulation

To figure out the image pattern and characteristics of the magnetic signal measured in our experiment, we performed magnetic signal simulation based on the magnetic dipole model and other physical principles. Firstly, we briefly describe the physical processes of the magnetic detection in our IMM. For a sample with magnetization distribution $\vec{M}(x,y)$, the generated magnetic field can be given by the dipole model as:

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi}\int\left[\frac{3\vec{M}\left(\vec{r}^{\,\prime}\right)\cdot\left(\vec{r}-\vec{r}^{\,\prime}\right)}{\left|\vec{r}-\vec{r}^{\,\prime}\right|^5}\left(\vec{r}-\vec{r}^{\,\prime}\right)-\frac{\vec{M}\left(\vec{r}^{\,\prime}\right)}{\left|\vec{r}-\vec{r}^{\,\prime}\right|^3}\right]d\vec{r}'$$

where μ_0 is the vacuum permeability. Under an external magnetic field $\vec{B}_0$ aligning to NV axis, magnetization possesses the same direction as the $\vec{B}_0$. The Zeeman shift of NV spin is proportional to the parallel component of the total magnetic field, that is:

$$B_{\parallel} = B_0 + \frac{\mu_0}{4\pi}\int\frac{3\left[\widehat{M}\cdot\left(\hat{r}-\hat{r}^{\,\prime}\right)\right]^2-1}{\left|\vec{r}-\vec{r}^{\,\prime}\right|^3}M\left(\vec{r}^{\,\prime}\right)d\vec{r}'$$

Meanwhile, the NV ensemble used in this work has a thickness of about 100 nm, and the distance between NVs is approximately 20 nm. Taking all of these into consideration, we can simulate the emitted fluorescence of NV ensemble in real space by:

$$C(x,y,f) = \sum_{i=1}^{N} c_0\left[1-\frac{4Cw^2}{4\left(f-\gamma B_{\parallel}(x,y,z(i))\right)^2+w^2}\right]$$

where c_0 is the fluorescence count rate, C and w are the continuous-wave (CW) contrast and HWFM of NV ensemble, respectively, f is the applied microwave frequency, and N denotes the number of NV centers along z axis. Assuming that all NV emitters share the same optical point spread function (PSF) PSF_o, which is given by our optical system, then the detected fluorescence image is given by the convolution:

$$C_{obj}(x,y,f) = C(x,y,f) * PSF_o$$

In addition, we need to discrete the real transverse plane into pixels, which is in accordance with camera acquirement mode, during the simulation. By analyzing fluorescence images acquired by the camera, we can eventually figure out the magnetic field image. Here we show some simulation results according to the physical principles above.

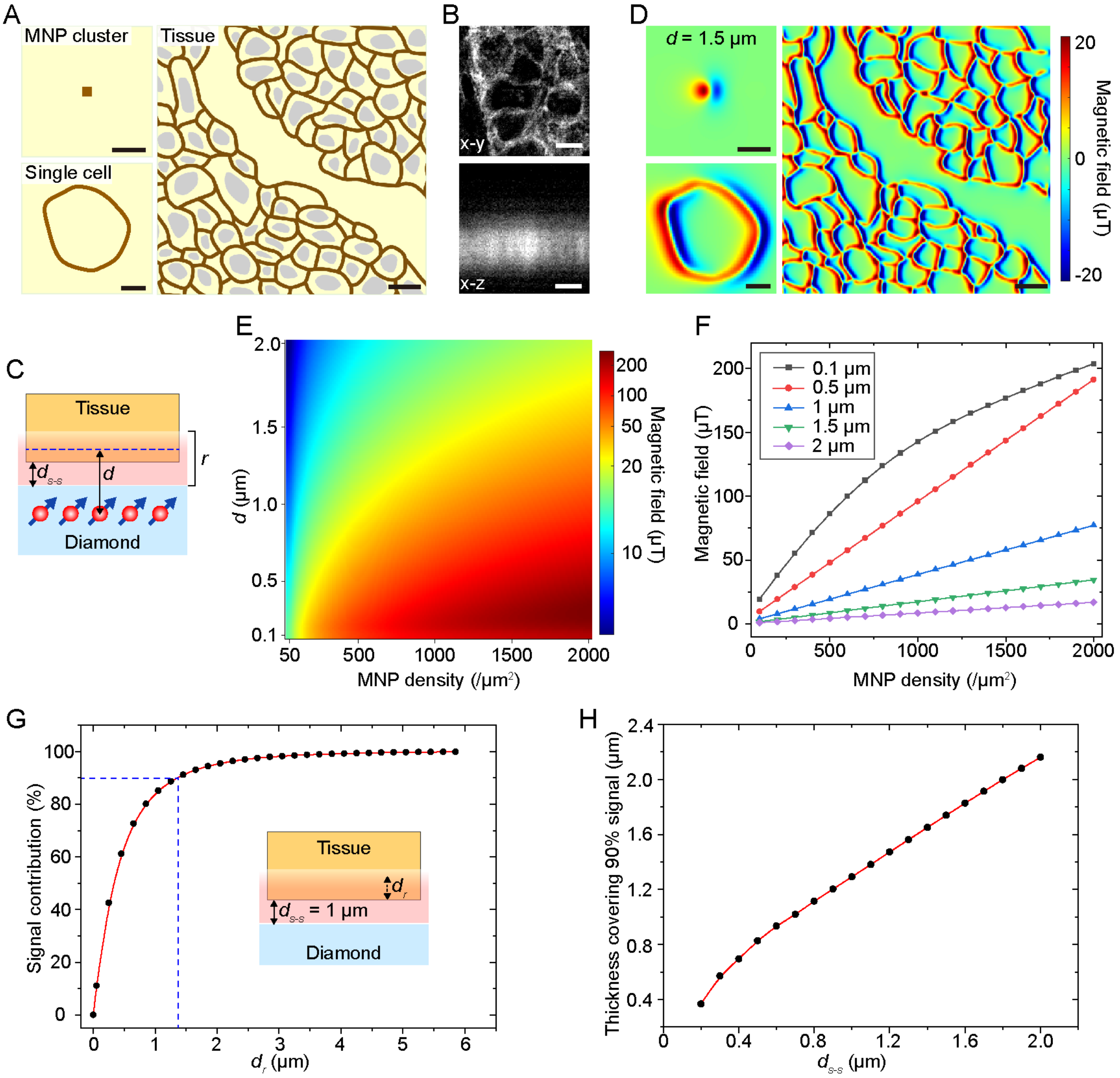

Fig. S2. Magnetic signal simulation. (*A*) Generated magnetic moment distributions of an MNP cluster, a single cell, and a tissue section with a maximum MNP density of 1,000/μm^2 for the magnetic signal simulation. (*B*) Confocal fluorescence images of a 20 nm MNP-labeled tissue section. The tissue was immuno-labeled with the bimodal magnetic-fluorescent label Cy3-MNP, as shown in Fig. S6*A*. The x-z scanning image displays the uniform distribution of MNPs and confirms the good permeability of 20 nm MNPs. Scale bars, 10 μm (x-y) and 2 μm (x-z). (*C*) Illustration of detecting an MNP-labeled tissue section by NV centers. $d_{s\text{-}s}$, the distance between the tissue surface and diamond surface; *d*, the equivalent distance between the sample and the sensors; *r*, the detection range in which the MNPs in the tissue generate 90% of the total signal. (*D*) Simulated magnetic images of *A* with a *d* of 1.5 μm. The red and blue lines mark the magnetic signal as two poles. The magnetic signal magnitude is between 20 μT and -20 μT. Scale bars, 5 μm (MNP cluster and single cell); 20 μm (tissue). (*E*) Magnetic field maximums of single MNP clusters as *A* with different MNP densities and *d*. (*F*) Representative curves extracted from *D* display the linear relation between the magnetic field intensity and the MNP density with a given *d*. The nonlinear phenomenon for the distance 0.1 μm may come from the competition of magnetic PSF and optical PSF as well as the dramatic change of magnetic field direction. (*G*) Signal contributions of layers with different accumulated thicknesses in a tissue section. With a $d_{s\text{-}s}$ of 1 μm, a ~1.3-μm-thick tissue layer contributed to 90% signal (illustrated by the blue dashed lines), which we named "magnetic slice". d_r, accumulated range. (*H*) Thickness of tissue layer contributing to 90% signal for different $d_{s\text{-}s}$.

Although we focus on magnetic imaging at the tissue level in this work, it is still beneficial for us to understand the magnetic signal pattern and magnitude range in the tissue if the magnetic images of single MNP cluster and single cell are simulated well. The simulation results showed that the magnetic signal distribution of a single MNP cluster resembled a dipole distribution with a 2-lobe pattern, while that of a single cell was the superposition of magnetic fields given by many MNP clusters (Fig. S2 *A* and *D*). Specifically, the MNP has a 20-nm iron oxide core and an about 6×10^{-16} emu magnetic moment. Furthermore, the simulated magnetic image of tumor tissue displayed a similar magnetic pattern as the single cell (Fig. S2 *A* and *D*), which matched well with the experimental magnetic images. Using the same physical model, we calculated the magnetic fields of single MNP clusters with different MNP densities and NV-sample distances, and the magnetic field maximums were extracted and illustrated (Fig. S2*E*), which reflected the inherent characteristics of the magnetic signal. We found that when the distance between the NV layer and the sample was larger than 0.5 μm, the magnetic field maximum was perfectly linear to MNP density (Fig. S2*F*). This property ensured the feasibility of immunomagnetic quantification and made magnetic imaging with a high dynamic range possible. Furthermore, we investigated the signal contribution of MNP-labeled tissue with different accumulated thicknesses and found that a shallow tissue layer close to the diamond dominated in the signal contribution (Fig. S2*G*). Meanwhile, the slice thickness contributing to 90% signal is of about a similar value as the distance between NV and tissue surface (Fig. S2*H*). These simulation results well interpret our experimental results and provide a theoretical model for the following deep learning.

3. Deep learning for magnetic signal reconstruction

Measured magnetic field images in our experiments provided information about the distribution and intensity of cancer biomarkers in tissues. However, before we can intuitively study the biomarkers, there are still several problems, such as signal overlapping and indirect linear dependence, unless the original distribution of biomarkers (magnetic moment images) can be reconstructed from the magnetic field images. For this reconstruction, commonly used reconstruction methods in magnetology are the least square fitting (1) and Fourier inversion reconstruction (2). They can resolve the magnetic moment distribution superiorly, provided that the relation between the magnetic field and the magnetic moment can be transformed to direct matrix multiplication or convolution operation, and the distance is given. However, due to the optical essence of ODMR based on NV centers and the unknown distance between the tissue sample and the NV sensors, we cannot simply use the above-mentioned matrix operations to describe the conversion between the magnetic moment and the magnetic field. Accordingly, traditional methods are unable to accomplish this magnetic reconstruction. In this work, we used a deep learning algorithm to realize the magnetic reconstruction. Then the effective distance between the tissue section and the NV-diamond sensor can be retrieved from the reconstruction result. The conversion and the distance coordinated the veracious reconstruction of magnetic images.

3.1 Deep learning network

Broadly speaking, the conversion based on deep learning belongs to a class of vision and graphics problems called image-to-image translation. One of the most powerful deep learning frameworks for this conversion is conditional Generative Adversarial Networks (cGANs). In this work, we trained our model using the pix2pixHD framework (3), a type of cGANs, to learn the above end-to-end conversion. The pix2pixHD network was composed of a generator G and multi-scale discriminators D_k and trained with image pairs (Fig. S3*A*). The coarse-to-fine generator consisting of a global network and a local enhancer produced magnetic moment images with high resolution by training magnetic field images with 2× downsampling size and original size, respectively (Fig. S3*B*). The generated magnetic moment images and their paired ground truth were downsampled and then used to train three discriminators with different receptive fields. The three discriminators possessed the same network structure (Fig. S3*C*). The training process was repeated for many epochs, e.g., 200. The ultimate goal of the generator G is to generate the most realistic magnetic moment image in the game with discriminator D when inputting a magnetic field image. The network information of different layers is listed below.

Theoretically, the objective function below depicts the competition between the generator and the discriminators, which combines the GAN loss and the feature map loss:

$$\min_{G}\left(\left(\max_{D_1,D_2,D_3}\sum_{k=1,2,3}\mathcal{L}_{GAN}(G,D_k)\right)+\lambda\sum_{k=1,2,3}\mathcal{L}_{FM}(G,D_k)\right)$$

where D_k represents the k^{th} discriminator at different scales, and λ reconciles the weights of the two terms. More specifically, the GAN loss $\mathcal{L}(G, D_k)$ is given by

$$\mathbb{E}_{(\mathbf{B},\mathbf{M})}[\log D_k(\mathbf{B},\mathbf{M})]+\mathbb{E}_{\mathbf{B}}\left[\log\left(1-D_k(\mathbf{B},G(\mathbf{B}))\right)\right]$$

and the feature map loss $\mathcal{L}_{FM}(G, D_k)$ is written as

$$\mathbb{E}_{(\mathbf{B},\mathbf{M})}\sum_{i=1}^{T}\frac{1}{N_i}\left[\left\|D_k^{(i)}(\mathbf{B},\mathbf{M})-D_k^{(i)}(\mathbf{B},G(\mathbf{B}))\right\|_1\right]$$

Where $^{(i)}$ denotes the i^{th}-layer feature extractor of D_k, **B** and **M** represent magnetic field image and magnetic moment image, respectively. The discriminators with different receptive fields try to maximize the loss $\mathcal{L}_{GAN}$, while the generator G attempts to minimize the total loss function. This supervised learning model endows pix2pixHD a strong ability to unravel the implicit relation between different image domains.

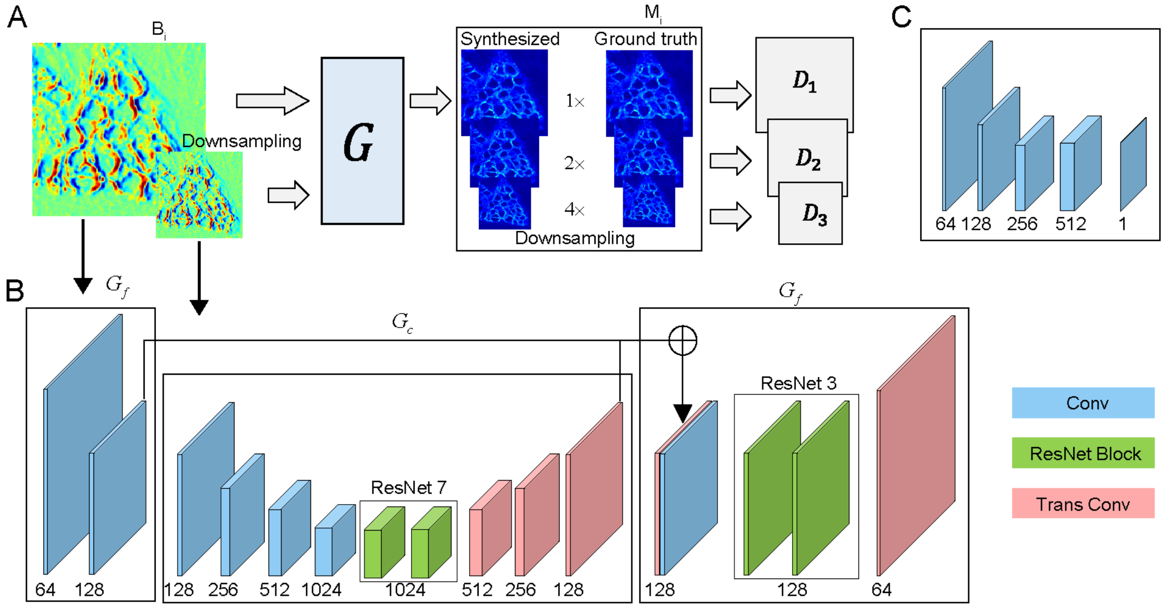

Fig. S3. Principle and framework of the deep learning model for magnetic moment reconstruction. Our deep learning model is based on the recently developed pix2pixHD network (3). (*A*) Training pipeline of magnetic moment reconstruction using simulated data. The training data contained thousands of image pairs {(B_i, M_i)}, in which magnetic moment images M_i were generated from experimental fluorescence images and magnetic field images B_i were simulated from M_i according to the physical principles in the previous section. During the training process, magnetic images were downsampled 2× and then put into the generator together. Synthesized and ground truth magnetic moment images were concatenated to train multi-scale discriminators with 1×, 2×, and 4× downsampling. (*B*) Architecture of generator. The network of generator consists of global generator G_c and local enhancer G_f, which learn coarse and fine structures, respectively. Network G_c was trained first, and then G_f was appended to G_c to be trained jointly. The input to ResNet in

G_f was the element-wise sum of feature maps of G_f and G_c. (*C*) Network architecture of discriminator. All three discriminators have the same architecture. The structures of all three network blocks (Conv, ResNet Block, Trans Conv) are listed in the next section **Training details**.

Due to the unavailability of enough experimental image pairs {(B$_i$, M$_i$)}, we generated training set and test set (See next section **Training details** for details) by simulating magnetic images according to the physical principles and parameters described above. We collected 442 fluorescence images of Cy3, labeled on MNPs (Fig. S6*A*), and IFM labeling various biomarkers, mainly membrane proteins, and slightly adjusted, modulated, and resized them as magnetic moment images M$_i$. And magnetic field images B$_i$ were simulated from the magnetic moment images M$_i$ with a given distance from 0.7 μm to 2.0 μm (the step size was 0.1 μm), and randomly corrupted by a uniformly distributed noise [0, 4 μT]. The natural random distributions of biomarkers in the simulated images are similar to those in the experimental data, which helps to eliminate the domain shift. All yielded images were processed to 6188 pseudo-color RGB image pairs (colormap: 16-bit jet, MATLAB) and put into pix2pixHD for subsequent training and testing (Fig. S3*A*). In order to evaluate our training model, we assessed three vital indicators: Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and the distance deviation on the test set, where SSIM and PSNR were calculated in MATLAB. The test set was randomly drawn out from simulated data with a known distance tag. As a result, we found that our model can simultaneously retrieve the distance and the magnetic moment distribution on the test set (Fig. S4 *A* and *B*). The distance was determined by the minimum location of magnetic field L2 loss as the formula $\min_{d} ||B - f_{M2B}(M, d)||^2$ showed, where the function f_{M2B} denoted the previous simulation process. Statistical analysis also supported this conclusion: SSIM = 0.933 ± 0.037, PSNR(dB) = 37.12 ± 4.09, and the distance retrieval deviation = 0.0828 ± 0.0945 μm; *n* = 1,188. The generalization ability of our trained model was further confirmed in experimental magnetic images by reconstructing their magnetic moment distribution and resolving the equivalent distances, e.g., the result of an experimental data, whose effective distance was calculated as 1.7 μm (Fig. S4 *C* and *D*). Combining the curve in Fig. S2*H*, we can estimate the NV-sample surface distance equaled to ~1 μm. Using this model, we can resolve the magnetic moment distributions of experimental magnetic field images and accomplish the quantification of biomarker expression.

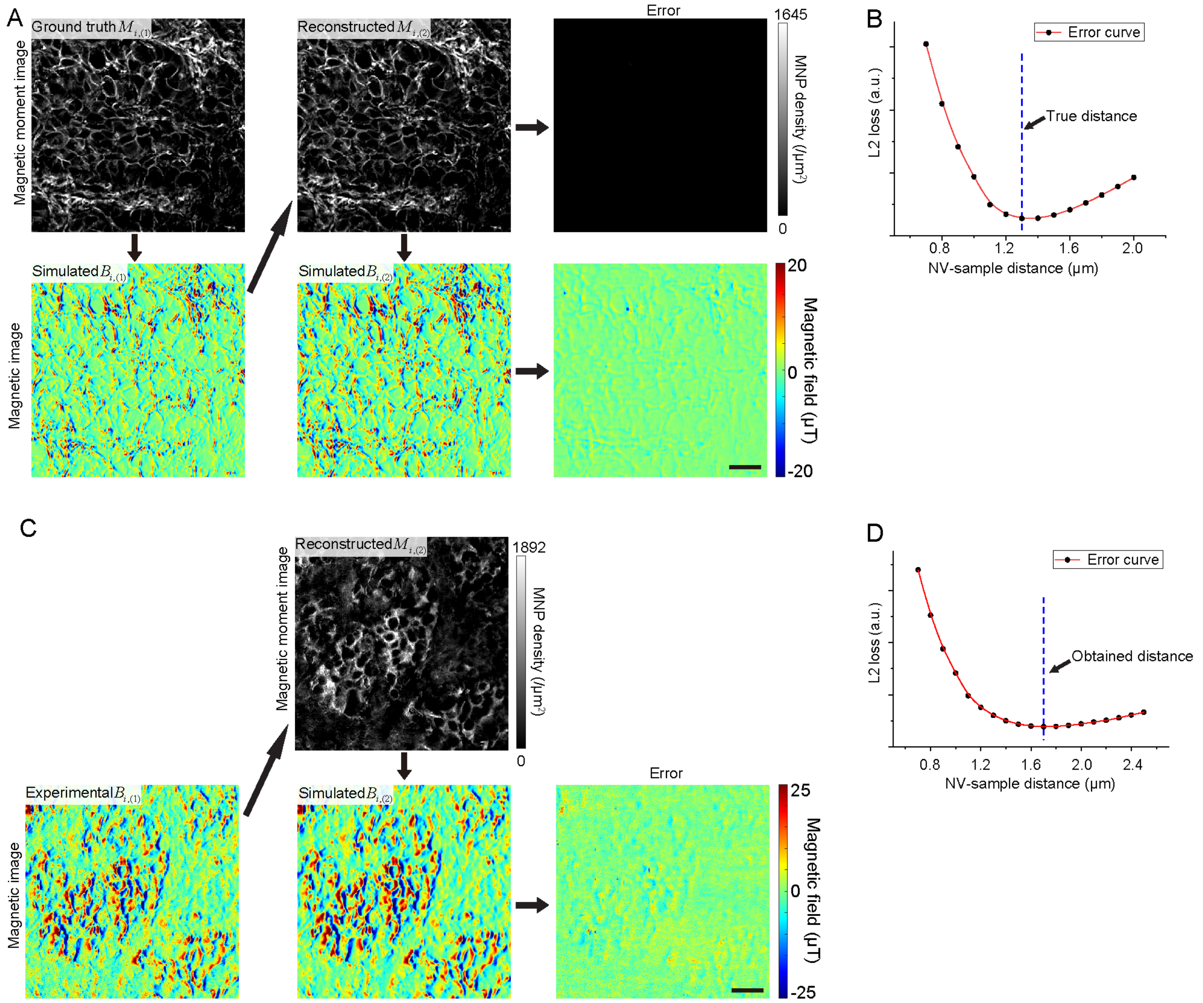

Fig. S4. Validation of pix2pixHD on test set and experimental result. (*A* and *B*) Magnetic moment reconstruction and distance retrieval on a test data. A magnetic moment image $M_{i,(1)}$ (Ground truth) was used to simulate the magnetic field image $B_{i,(1)}$ with the ground truth distance $d = 1.3$ µm. Then we tested our trained model by reconstructing magnetic moment image $M_{i,(2)}$ using the simulated magnetic field image $B_{i,(1)}$. Here all magnetic moment images (RGB type) were processed to gray images based on the colormap (16 bit jet, MATLAB), and displayed as MNP density images with a single MNP moment of 6×10^{-16} *emu*. Afterwards, we simulated magnetic field images $B_{i,(2)}$ with $M_{i,(1)}$ and different distances, subtracted previously simulated $B_{i,(1)}$, and retrieved the distance by determining the minimum location of the L2 loss curve $\| B_{i,(1)} - B_{i,(2)} \|^2$ in *B*. The resolved distance matched the ground truth distance represented by the blue dashed line. These results showed that our trained model was able to resolve the magnetic moment image and the distance simultaneously. (*C* and *D*), Magnetic moment reconstruction and distance retrieval on an experimental data. We analyzed the experimental data in the order of $B_{i,(1)} \rightarrow M_{i,(2)} \rightarrow B_{i,(2)} \rightarrow \| B_{i,(1)} - B_{i,(2)} \|^2$ and retrieved the distance 1.7 µm, which was also determined using L2 loss curve. The existence of a minimum in L2 loss curve confirmed the generalization ability of our trained model for simulated and experimental data. Scale bars, 50 µm.

3.2 Training details

We trained the pix2pixHD networks using Adam solver from scratch on a GPU cluster with 4 × NVIDIA V100 (16 G) in a supercomputing system (University of Science and Technology of China). The whole training consisted of two periods: constant learning rate 0.0002 for 200 epochs and linearly decayed rate to zero over the next 200 epochs. All weights were initialized from a Gaussian distribution with mean 0 and SD 0.02. Total 5,000 image pairs $\{(B_i, M_i)\}$ with an image size of 512 × 512 pixels were used as the training set for training and 1188 image pairs were used as the test set for validation and distance deviation analysis. For the generator, the Conv block contains a 3 × 3 Convolutional-InstanceNorm-ReLU layer with k filters. The ResNet Block denotes a residual block that contains two 3 × 3 convolutional layers with the same number of k filters, and the ResNet N means N ResNet blocks used. The Trans Conv block is composed of 3 × 3 fractional-strided-Convolution-InstanceNorm-ReLU layer with k filters, and stride 1/2. For discriminators, the Conv block contains a 4 × 4 Convolutional-InstanceNorm-LeakyReLU layer with k filters, and stride 2. We used leaky ReLUs with slope 0.2. The filter number k in all layers is provided in Fig. S3 *B* and *C*.

4. Magnetic quantification error analysis

To evaluate the quantification accuracy of our method, we further analyzed magnetic reconstruction and measurement errors. Based on the fact that our training data were simulated from experimental fluorescence images with naturally random distributions of biomarkers, we decomposed quantification errors into three parts: (1) Deep learning network's original reconstruction error on the test set; (2) Magnetic quantification error purely due to the magnetic field measurement noise; (3) The domain shift from the training data to the experimental data. All these types of errors are analyzed in detail as follows.

Here we focused on the reconstructed magnetic moment image instead of the retrieved distance because they actually were correlated by the convolution operation. When magnetic field images in the test set were put into our trained deep learning network, it output reconstructed magnetic moment images which slightly deviated from the ground truth. The magnetic moment residuals were analyzed and characterized by figuring out the standard deviation σ_0 of the residual image (Fig. S5*A*). For 500 randomly selected images in the test set, we took the average and finally the average original reconstruction error of the deep learning network was obtained as $\overline{\sigma_0} = 39.388 \pm 21.453$ MNPs/μm^2 (±1σ, 65% confidence interval). This error represented the original reconstruction accuracy of our deep learning network.

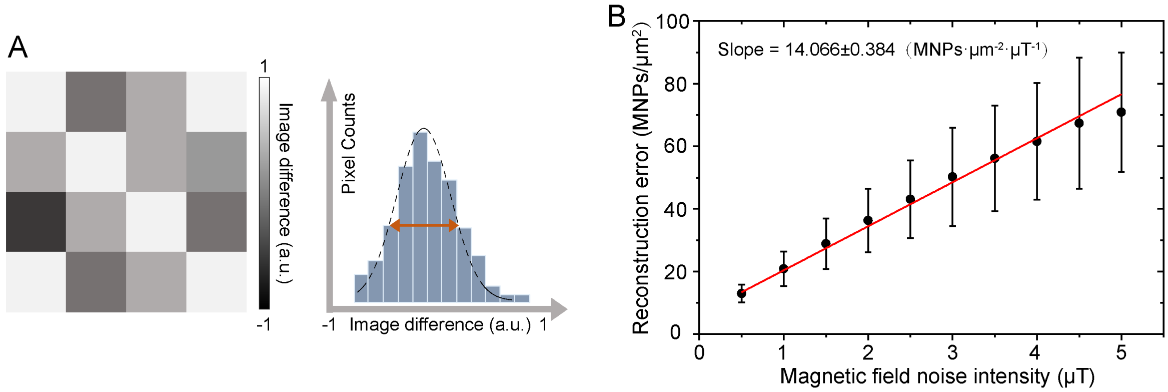

Fig. S5. Reconstruction error analysis for deep learning algorithm. (*A*) Evaluation of image residuals between the reconstructed image and the ground truth. Two-dimensional residual image was reshaped to the histogram and fitted with Gaussian distribution. The illustrated FWHM by double arrow indicated average reconstruction error between the reconstructed image and the ground truth. (*B*) Noise-dependent reconstruction error on experimental data. The Gaussian noise was added to the experimental magnetic field image. By comparing the difference between the reconstructed magnetic moment images with and without noise, we can estimate the effect of magnetic measurement noise on the quantification accuracy of our method.

However, during the magnetic reconstructing of experimental data, measurement noise can also cause quantification error. In order to evaluate the noise-dependent reconstruction error, we evaluated the quantification error caused by the magnetic field measurement noise. We selected 33 magnetic field images from our experimental data and corrupted them with varying degrees of Gaussian noises, whose intensities were denoted as σ_{noise}. Assuming that in a small neighborhood of the experimental image without noise corruption, the quantification error is linear to the corrupting noise intensity. By computing the average quantification error similarly defined above, we found the error was indeed linear to the noise intensity, provided that the intensity was much smaller than the normal magnetic field. The linear dependence relation was confirmed and fitted. The slope was obtained as $k = 14.066 \pm 0.384$ MNPs·μm^{-2}·μT^{-1} (Fig. S5*B*). Furthermore, we analyzed some blank sample data and figured out the magnetic measurement noise $\bar{\sigma}_{noise} = 1.192 \pm 0.002$ μT. Combining the slope and the noise level, the noise-dependent magnetic reconstruction error was estimated as $\sigma_N = k\cdot \bar{\sigma}_{noise} = 16.767 \pm 0.458$ MNPs/μm^2.

Aside from the effect of measurement noise, the domain shift between the training data and experimental data might lead to considerable quantification inaccuracy. Then we employed a self-cycled comparison to delineate the error caused by the domain shift due to the lack of ground truth for experimental data. The procedure contains: (a). Reconstruct magnetic moment images from magnetic field images; (b). Simulate magnetic field images with different distances; (c). Find the minimum of L2 loss for magnetic field images as Fig. S4*D*; (d) Evaluate the magnetic field image difference in (c) and estimate the domain shift error σ_{ds}. Using the same 33 experimental magnetic field images, we figured out $\sigma_B = 3.567 \pm 1.612$ μT, then the reconstruction error caused by the domain shift was $\sigma_{ds} = k\cdot\sigma_B = 50.173 \pm 22.716$ MNPs/μm^2. The difference between σ_{ds} and $\bar{\sigma}_0$ estimated above indicated the domain shift was small, which confirmed the generalization ability of our trained deep learning network.

Based on the error analysis above, the total quantification error in our method is $\sigma_{exp} = (\sigma^2_{ds}+\sigma^2_N)^{1/2} = 52.901 \pm 21.545$ MNPs/μm^2 for a single pixel. Compared with the typical MNP density of 1000 MNPs/μm^2 in the experimental data, the quantification error was ±5.29%. Moreover, the quantification error at the single-cell level (with a diameter of 10 μm and the typical MNP density is 400 MNPs/μm^2) can be further eliminated to approximately ±7.5 MNPs/μm^2 or ±1.875%, which confirmed the magnetic quantification accuracy of our method.

5. Immunomagnetic microscopy of cultured cells

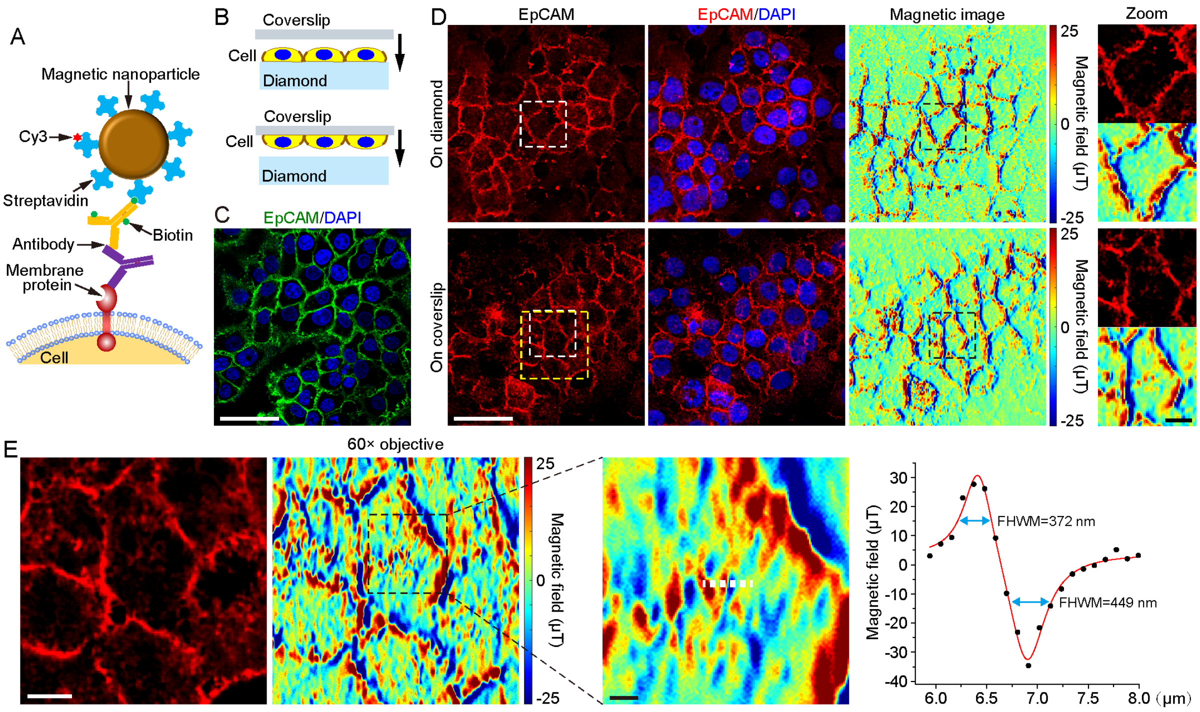

Fig. S6. Immunomagnetic microscopy of cultured cells. (*A*) Schematic of the immunomagnetic labeling. The primary antibody recognizes the biomarker, a membrane protein, and the biotinylated secondary antibody IgG binds the primary antibody. Then the streptavidin-conjugated iron oxide nanoparticle couples with the biotinylated IgG via the biotin. Optionally, the fluorescent probe Cy3-biotin (See Methods for detail) labels the cell via unbound biotin-binding domains of streptavidin on MNPs, forming a bimodal magnetic-fluorescent label Cy3-MNP. (*B*) Schematic of the treatment of MNP-labeled cells before the magnetic examination. Cells cultured on glass coverslip and diamond were embedded using UV curing adhesive, forming sandwich structures with diamond and coverslip, respectively. The pressing brought the sample and the sensor close, and the adhesive fixed the connection. (*C*) Fluorescence image of NCI-H292 cells. The cells were processed by using routine immunofluorescence microscopy. Alexa Fluor 488 in the green channel represents the epithelial cell adhesion molecules (EpCAM). DAPI in the blue channel stained cell nuclei. (*D*) Images of NCI-H292 cells cultured on coverslip and diamond, respectively. The cells were magnetic-labeled, targeting EpCAM by the immunomagnetic labeling protocol shown in *A*. Cy3 in the red channel represents the original location of MNPs. DAPI stained cell nuclei. Magnified single-cell images of regions of interest (ROI) in the dotted boxes showed representative subcellular magnetic field patterns. The magnetic images with ~1 μm resolution displayed high contrast and strong signal along the plasma membrane, which was consistent with the distributions of EpCAM in c and Cy3 labeled on MNPs. (*E*) Magnetic image with a 60× objective of the ROI shown in the yellow dotted box of panel *D*. The magnetic imaging realized a high spatial resolution of approximately 400 nm. Two full widths at half maximum

(FWHM) of a magnetic dipole were shown in the graph. Scale bars, 50 µm (*C* and *D*); 10 µm (*E* and zoom in *D*); 2 µm (zoom in *E*).

6. Fluorescence characterization of human lung cancer tissues

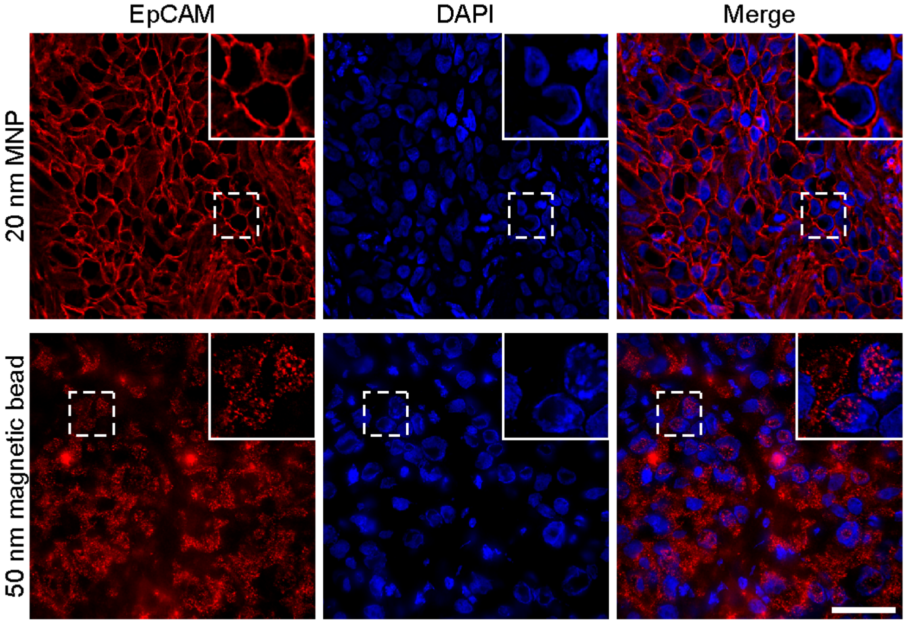

Fig. S7. The smaller magnetic nanoparticles have better specificity than the larger ones for the immunomagnetic labeling. The membrane protein EpCAM in lung tumor tissues were immunomagnetically labeled with the bimodal label Cy3-MNP shown in Fig. S6*A*, and imaged in the fluorescence microscope. 20 nm MNP-labeled cells displayed membrane contours without apparent nonspecific signals. In contrast, although the larger one has a theoretically stronger magnetic field signal, the labeling of 50 nm magnetic beads produced disordered punctum signals in cells, and membrane contours were almost invisible. Possibly because the poor permeability of larger particles in the tissue limited their performance. DAPI stained cell nuclei. Scale bar, 50 μm.

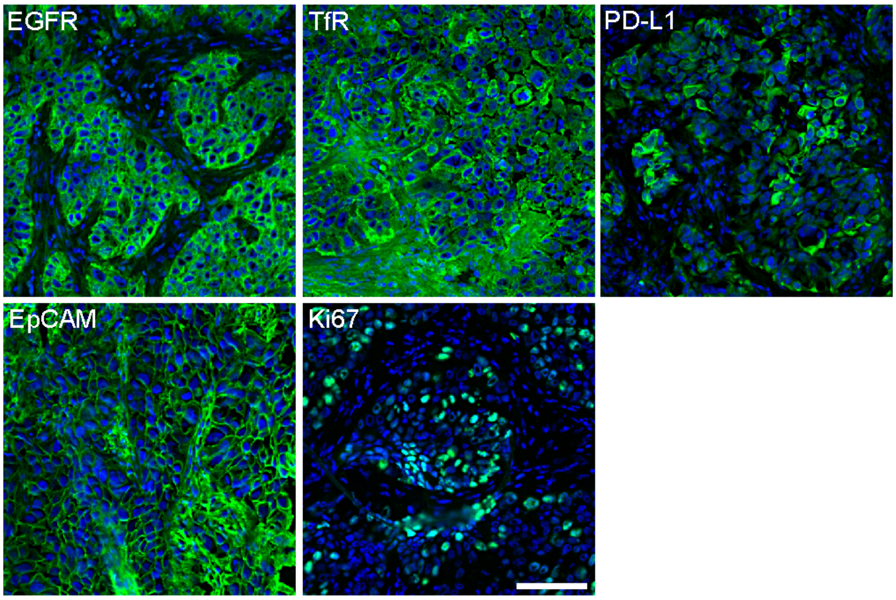

Fig. S8. IFM images of a variety of cancer biomarkers in human lung cancer tissues. The proteins were immuno-labeled by routine immunofluorescence procedure. Alexa Fluor 488 labeled biomarkers and DAPI stained cell nuclei. Scale bar, 100 μm.

7. IMM has excellent signal stability

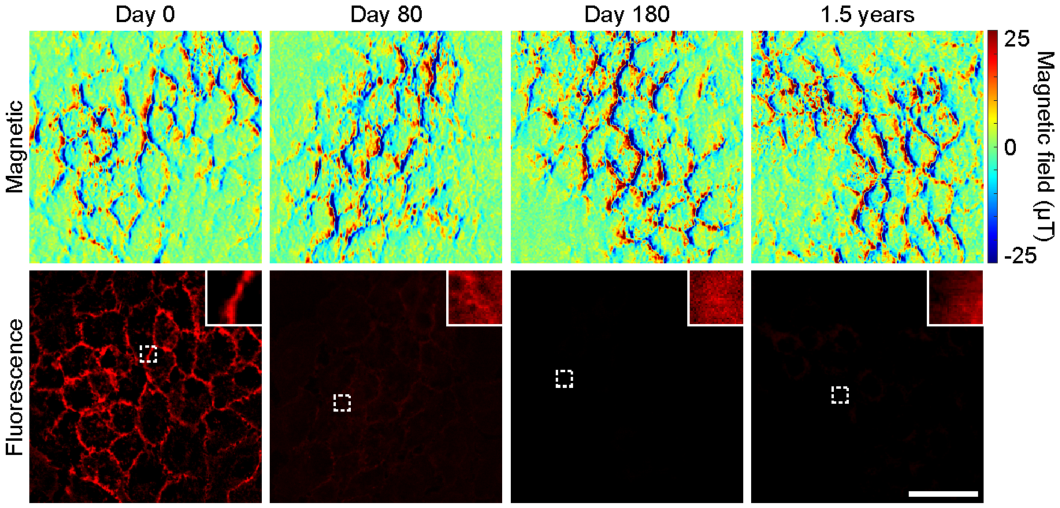

Fig. S9. IMM has excellent signal stability. Comparison of the stability of fluorescence and magnetic signals. NCI-H292 cells cultured on a coverslip were immunomagnetically labeled with the bimodal label Cy3-MNP. The coverslip with cells was cut into several parts. One of them was examined in the magnetic microscope as the time point day 0. The remaining samples were transferred under the illumination of a fluorescent lamp with an optical power density of less than 1 mW/cm^2. After the time point day 180, the remaining samples were transferred under ambient conditions without the fluorescent lamp. In the fresh sample, the images exhibited ideal signals, both magnetic and fluorescence. However, under the lamp illumination, the Cy3 fluorescence was quenched continuously until it completely disappeared. In contrast, the magnetic signal exhibited excellent stability, even after over a year and a half. Although magnetic signals from the four groups look slightly different, their intensity and distribution are almost the same. Scale bar, 50 μm.

8. IMM in liver tumor tissues

The autofluorescence in liver tumor tissues significantly affects fluorescence imaging, whereas it just has little impact on magnetic field measurement. Two aspects ensure this state. The first one is the technical issue. In addition to using an optical filter matching the optical spectrum of NV center (Materials and Methods), in this study, the 100-nm NV layer contributes to the high fluorescence count, which is about one order of magnitude brighter than the autofluorescence in the NV optical spectrum channel. The second and major aspect is the physical principle issue. IMM is based on the frequency shift measurement manner (Figure 1*B*, Materials and Methods), and the autofluorescence will only slightly decrease the contrast of NV centers' CW spectrum but not affect the peak position. As shown in Fig. S10, there is no noticeable difference between the spectral contrasts of NV centers with or without liver tumor tissue. These results were confirmed by the low noise in the background magnetic image of liver tumor tissue in Figure 3*D*.

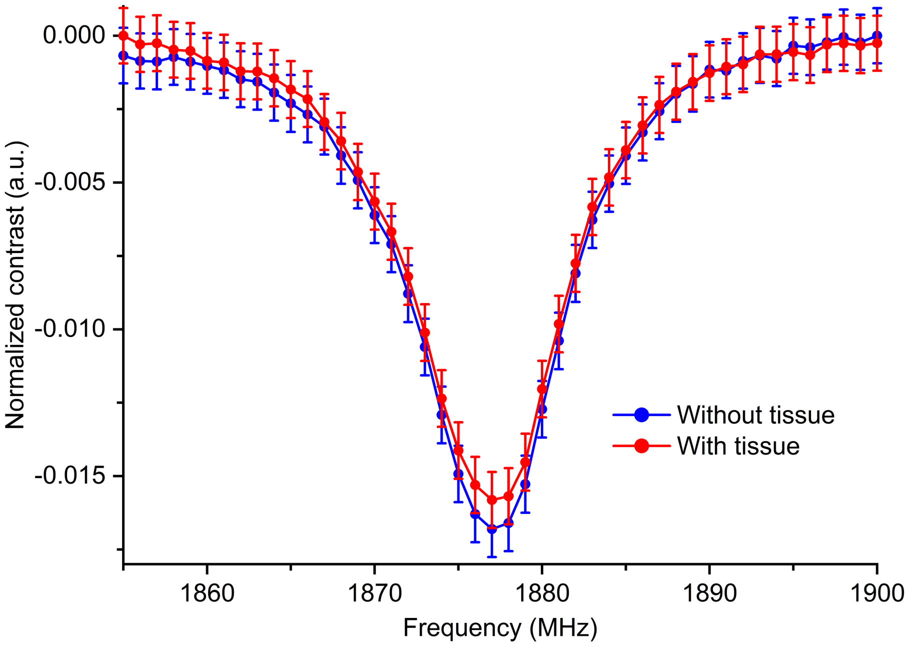

Fig. S10. Comparison of spectral contrasts of NV centers with or without liver tumor tissue. There is no noticeable difference between the two spectral contrasts.

Besides cryosections, we carried out the immunomagnetic imaging in a paraffin-embedded mouse normal liver sample in our hand (Fig. S11). Due to the strong background fluorescence signal of liver samples, histone H2B with a characteristic subcellular distribution was selected and labeled. The results show that the immunofluorescence signal is seriously affected by the autofluorescence in the tissue and the image contrast is poor (Fig. S11*A*), while IMM obtained high contrast images (Fig. S11*B*).

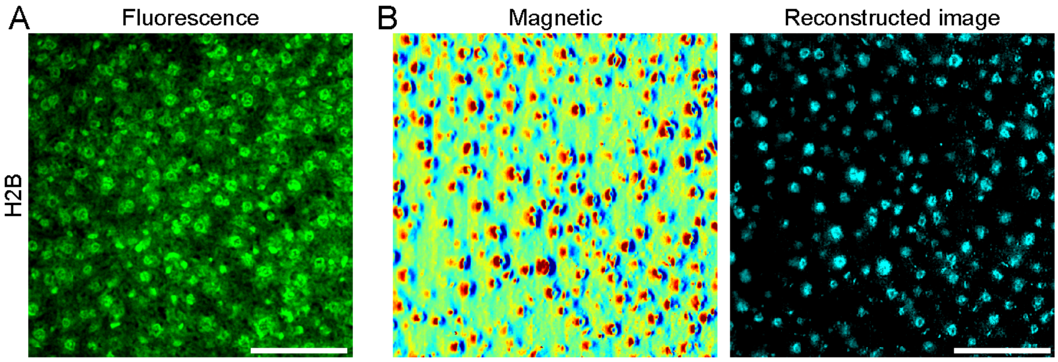

Fig. S11. IMM in paraffin-embedded mouse liver tissue. (*A*) Immunofluorescence image of H2B in a paraffin-embedded mouse liver tissue section. Alexa Fluor 488 in the green channel was labeled via the routine immunofluorescence procedure. There are obvious autofluorescence signals in the cytoplasm besides the specific signal in the nuclei. (*B*) IMM images of H2B in another liver tissue section. Scale bars, 100 µm.

9. Fluorescence images and CW spectrums in correlated imaging

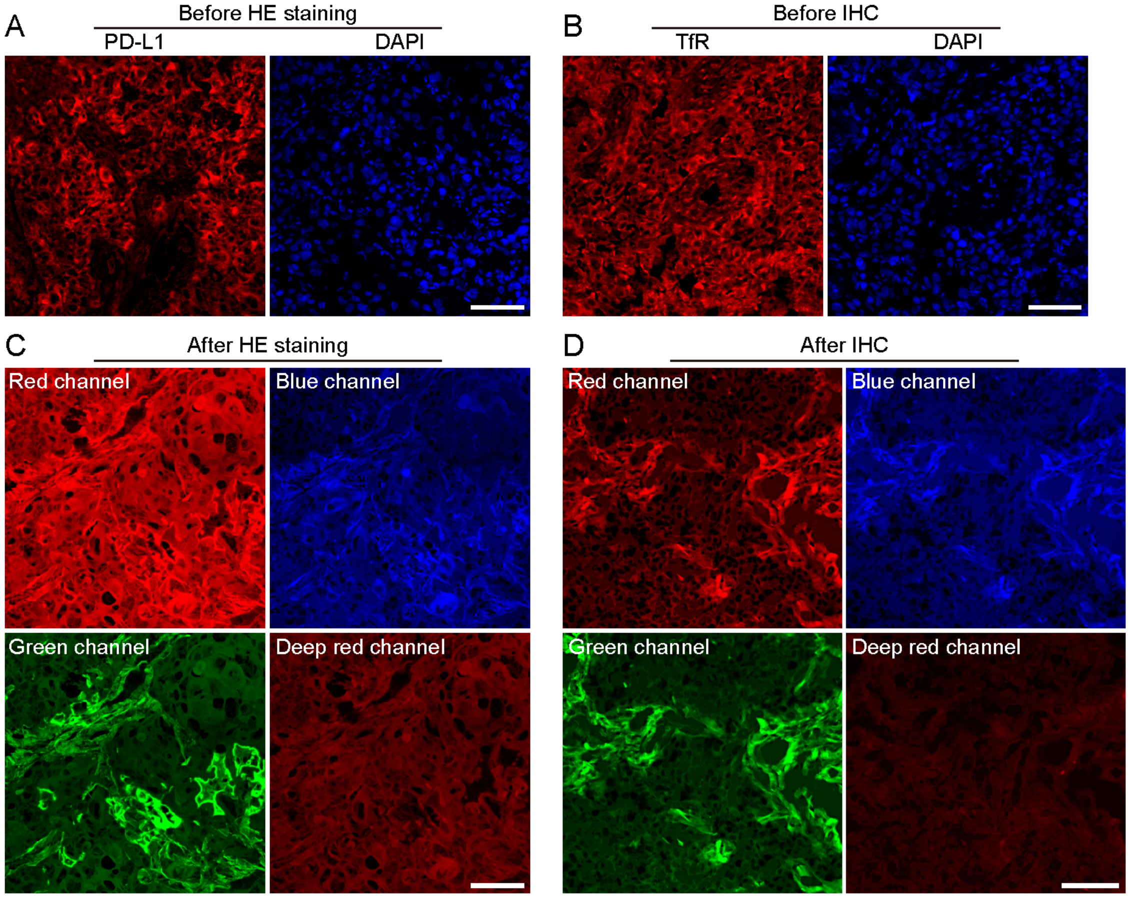

Fig. S12. HE and IHC staining destroyed the fluorescence in tissues and produced autofluorescence. (*A* and *B*) Fluorescence images of immuno-labeled tissues before HE or IHC staining. Cy3 was the component of bimodal label Cy3-MNP, and DAPI stained cell nuclei. (*C* and *D*) Fluorescence mages after HE and IHC staining in the same areas of Fig. 6 *A* and *B*, respectively. The staining produced obvious autofluorescence signals in nearly all fluorescence channels and seriously affected routine fluorescence staining. The deep red channel represents the NV-fluorescence channel. Other fluorescence channels are described in Methods. Scale bars, 100 μm.

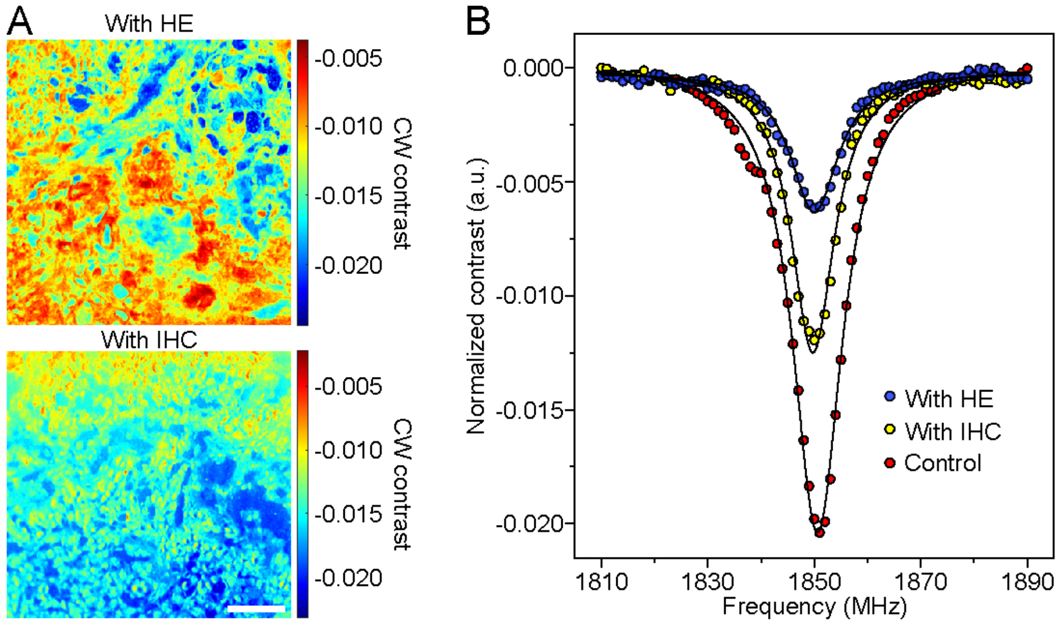

Fig. S13. Maps of normalized spectral contrasts of NVs related to Fig. 6. (*A*) Contrasts of NVs' CW spectrums were significantly reduced by HE or IHC staining. The distribution of different contrasts is related to the distribution of dyes. Scale bar, 100 μm. (*B*) Representative CW spectrums of NVs in control, HE, and IHC stained tissues, respectively.

10. SI References

1. B. P. Weiss, E. A. Lima, L. E. Fong, F. J. Baudenbacher, Paleomagnetic analysis using SQUID microscopy. *Journal of Geophysical Research: Solid Earth* **112** (2007).
2. E. A. Lima, B. P. Weiss, L. Baratchart, D. P. Hardin, E. B. Saff, Fast inversion of magnetic field maps of unidirectional planar geological magnetization. *Journal of Geophysical Research: Solid Earth* **118**, 2723-2752 (2013).
3. T.-C. Wang *et al.*, High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8798–8807 (2018).