Supporting Information

An array-based approach to determine different subtype and differentiation of non-small cell lung cancer

Chao Li[1*], Yucai Yang[2*], Luming Wei[1], Xiaoying Wang[3], Zhaoxia Wang[2], Yongmei Yin[3], and Genxi Li[1]

1. Department of Biochemistry and State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing 210093, China
2. Department of Oncology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing 210011, China
3. Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China

[*] These two authors contributed equally.

Corresponding authors: Genxi Li, Fax: +86-25-83592510, E-mail: genxili@nju.edu.cn; Zhaoxia Wang, Fax: +86-25-58509994, E-mail: zhaoxiawang88@hotmail.com; Yongmei Yin, Fax: +86-25-83710040, E-mail: ym.yin@hotmail.com.

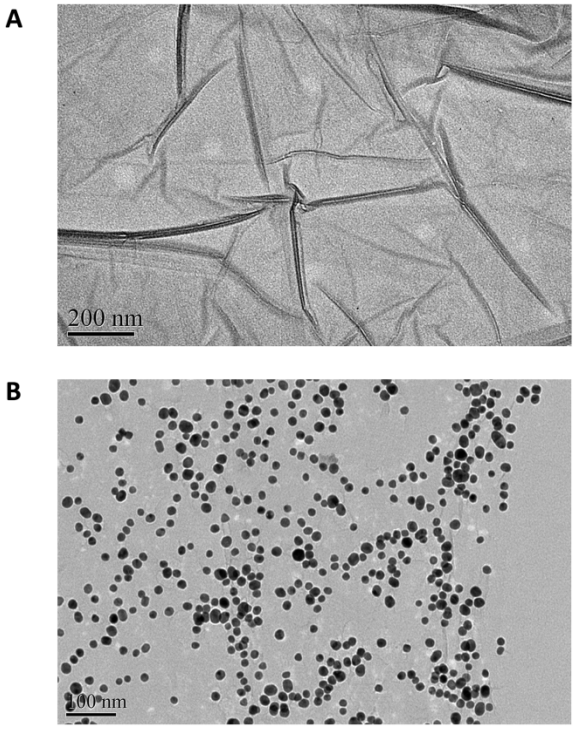

Figure. S1. Transmission electron microscopy (TEM) images of GO (A) before and (B) after incubation of ssDNA-AuNPs.

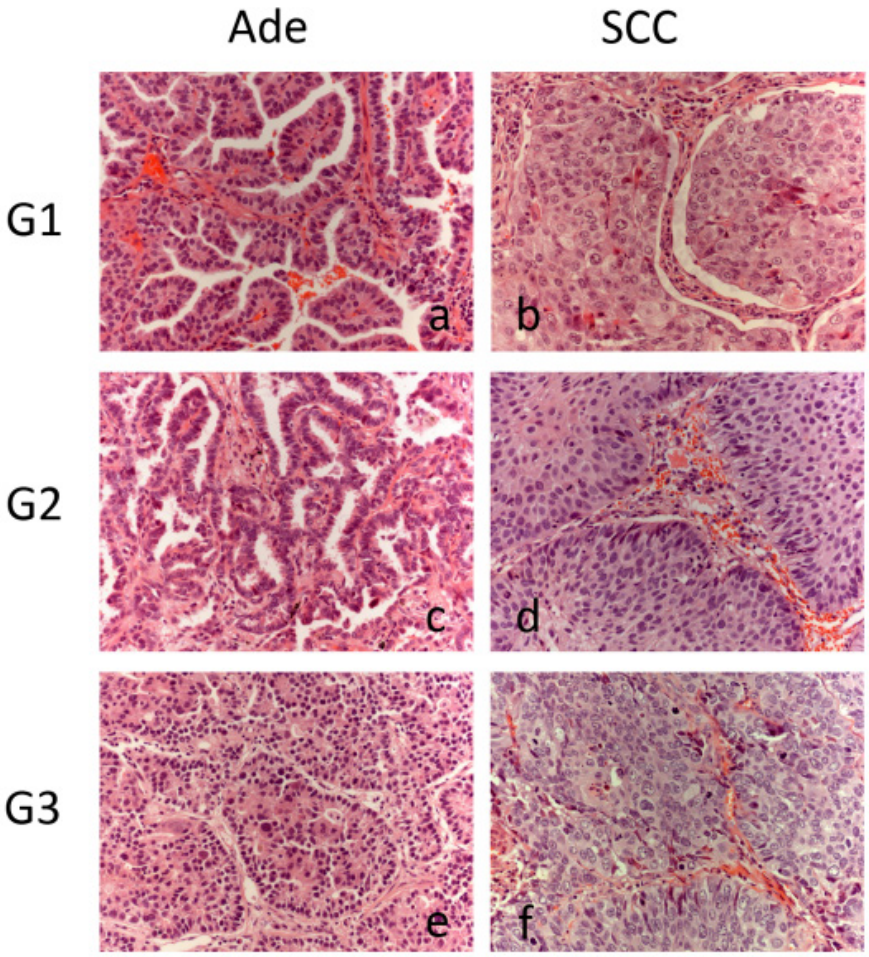

Figure. S2. Histological examination images for different subtypes and grades of NSCLC patients. (a, c, e) Histological images of G1 (well-differentiated), G2 (moderately-differentiated), and G3 (poorly-differentiated) of Ade patients. (b, d,f) Histological images of G1 (well-differentiated), G2 (moderately-differentiated), and G3 (poorly-differentiated) of SCC patients.

Table S1. Stern−Volmer binding constants (K_{s-v}) and effective ssDNA-AuNPs footprints (β) between GO and various ssDNA-AuNPs (NP1–NP6) as determined from absorbance titration.

Nanoparticle	K_{s-v} (μg^{-1})	β (nm^2)
NP1	15 ± 0.72	3887
NP2	11.88 ± 1.21	3989
NP3	1.11 ± 0.065	8745
NP4	6.19 ± 0.503	4440
NP5	1.81 ± 0.007	6728
NP6	0.71 ± 0.034	11661

Table S2. The eight analyte proteins used in sensing, and their properties of interest.

protein	MW (kDa)	pI
Bovine serum albumin	66.3	4.8
Lysozyme	14.4	11.0
Cytochrome c	12.3	10.7
Hemoglobin	64.5	6.8
Myoglobin	17	7.2
Horseradish peroxidase	40	6.5
Transferrin	76	5.9
Thrombin	36.7	7.05

Estimation of the binding constants from absorbance titration

The complexation of nanoparticles with GO could be expressed by equation S1, which assuming that GO has independent binding sites with DNA modified nanoparticles.

$$K_s = \frac{[L-S]}{[L][S]} \qquad \text{(S1)}$$

Where [L] and [S] are free nanoparticle and binding site concentrations respectively, and [L-S] is the concentration of the complex. With consideration that $[L_0]$ = [L] + [L-S], where $[L_0]$ denote the initial concentrations of ligand, and the number of binding sites can be described by [S] =n * [GO], the following relationship can be derived [1].

$$\frac{[L]}{[L_0]} = (1 + n * K_s[GO])^{-1}$$

Because of this equation is similar to an inverted Stern-Volmer equation, we can consider Stern-Volmer constant $K_{s-v} = n * K_s$.

$$\frac{[L]}{[L_0]} = (1 + K_{s-v}[GO])^{-1}$$

Given that graphene is absolutely dispersed and all the atoms other than carbon in the composition of graphene are neglected, we can easily calculate a theoretical effective footprint value β nm^2 of ssDNA-AuNPs as follows [2,3]:

$$\frac{1}{\beta} = \frac{\Delta A * N_A * C_0 * V}{\frac{m_{GO}}{D_g}}$$

C_0 is the initial concentration of ssDNA-AuNPs, ΔA is the absorbance change, N_A is Avogadro's number, V is the volume (100 μL), m_{GO} is the weight of graphene oxide, and D_g is the density of graphene per unit area (0.77 mg /m^2).

Table S3. Data matrix of absorbance response patterns against eight proteins from 0.5 nM sensing experiments in buffer.

Protein	NP1	NP2	NP3	NP4	NP5
1	0.104	0.366	-0.28	0.064	0.146
1	0.079	0.372	-0.296	0.088	0.136
1	0.089	0.388	-0.308	0.086	0.130
1	0.110	0.354	-0.312	0.075	0.154
1	0.094	0.396	-0.273	0.072	0.138
1	0.111	0.387	-0.297	0.083	0.146
2	0.278	0.122	-0.224	0.078	0.214
2	0.283	0.190	-0.212	0.089	0.186
2	0.298	0.152	-0.234	0.077	0.228
2	0.262	0.165	-0.256	0.072	0.198
2	0.267	0.178	-0.240	0.080	0.200
2	0.286	0.148	-0.228	0.084	0.194
3	0.118	0.348	-0.092	0.212	0.218
3	0.105	0.346	0.002	0.187	0.194
3	0.089	0.356	-0.102	0.201	0.200
3	0.092	0.338	-0.890	0.208	0.206
3	0.108	0.344	-0.045	0.194	0.200
3	0.112	0.345	-0.089	0.224	0.205
4	0.103	0.400	-0.024	0.150	0.176
4	0.113	0.428	-0.055	0.163	0.180
4	0.091	0.412	-0.062	0.144	0.182
4	0.087	0.420	-0.07	0.138	0.182
4	0.103	0.418	-0.044	0.148	0.170
4	0.105	0.417	-0.033	0.154	0.178
5	0.230	0.260	-0.320	0.116	0.184
5	0.201	0.267	-0.324	0.123	0.170
5	0.217	0.278	-0.336	0.134	0.160
5	0.234	0.268	-0.308	0.107	0.152
5	0.217	0.250	-0.304	0.118	0.142
5	0.239	0.262	-0.317	0.125	0.165
6	-0.023	0.390	-0.280	0.172	0.216
6	-0.014	0.420	-0.200	0.189	0.210
6	-0.029	0.406	-0.222	0.190	0.190
6	-0.058	0.400	-0.232	0.197	0.200
6	-0.078	0.380	-0.242	0.174	0.208
6	-0.002	0.396	-0.218	0.187	0.198
7	0.068	0.314	-0.274	-0.026	0.160
7	0.043	0.306	-0.266	0.009	0.144
7	0.106	0.306	-0.270	-0.012	0.150
7	0.088	0.310	-0.282	0.003	0.142
7	0.076	0.300	-0.272	0.010	0.128
7	0.100	0.305	-0.276	-0.013	0.148
8	0.224	0.432	-0.052	0.008	0.022
8	0.189	0.438	-0.028	-0.013	0.036
8	0.193	0.422	-0.048	-0.008	0.032
8	0.201	0.430	-0.056	0.013	-0.006
8	0.242	0.419	-0.032	0.008	0.038
8	0.198	0.424	-0.062	-0.003	0.022

1. Bovine serum albumin 2. Lysozyme 3. Cytochrome c 4. Hemoglobin 5. Myoglobin 6. Horseradish peroxidase 7. Transferrin 8. Thrombin

Table S4. The Jackknifed classification matrix showing the contribution of each nanoparticle in the differentiation against eight proteins.

NP1	NP2	NP3	NP4	NP5	%
					62
					79
					30
					65
					40
					100

Table S5. Detection and identification of 25 unknown proteins in buffer, the concentrations of prepared unknown protein samples are 0.5 nM, which is the same with that in training matrix (table S2).

Samples	NP1	NP2	NP3	NP4	NP5	Identification	Verification
1	0.063	0.316	-0.265	0.012	0.128	7	7
2	0.302	0.155	-0.197	0.092	0.206	2	2
3	0.295	0.189	-0.243	0.103	0.238	2	2
4	0.227	0.421	-0.074	0.000	0.044	8	8
5	0.229	0.280	-0.288	0.115	0.170	5	5
6	0.112	0.392	-0.266	0.091	0.145	1	1
7	0.089	0.351	-0.107	0.206	0.213	3	3
8	0.106	0.434	-0.033	0.174	0.182	4	4
9	0.111	0.427	-0.105	0.158	0.179	4	4
10	0.225	0.278	-0.305	0.117	0.174	5	5
11	0.122	0.400	-0.301	0.088	0.126	1	1
12	0.034	0.387	-0.269	0.102	0.172	4	1 fail
13	0.023	0.285	-0.255	-0.018	0.117	7	7
14	0.113	0.310	-0.280	0.104	0.119	1	1
15	0.236	0.422	-0.088	0.006	0.028	8	8
16	0.114	0.410	-0.064	0.157	0.171	4	4
17	0.110	0.317	0.046	0.195	0.214	3	3
18	0.251	0.184	-0.260	0.096	0.235	2	2
19	0.075	0.360	0.022	0.201	0.217	3	3
20	-0.069	0.409	-0.186	0.212	0.200	6	6
21	0.241	0.269	-0.344	0.095	0.166	5	5
22	0.004	0.397	-0.235	0.182	0.202	6	6
23	0.123	0.356	-0.337	0.103	0.160	1	1
24	0.102	0.412	-0.025	0.128	0.194	4	4
25	0.287	0.146	-0.219	0.067	0.230	2	2

1. Bovine serum albumin 2. Lysozyme 3. Cytochrome c 4. Hemoglobin 5. Myoglobin 6. Horseradish peroxidase 7. Transferrin 8. Thrombin

Table S6. Data matrix of absorbance response patterns against five proteins from 5 nM sensing experiments in serum.

Protein	NP1	NP2	NP3	NP4	NP5
1	0.213	0.422	0.152	0.353	0.340
1	0.187	0.434	0.189	0.338	0.336
1	0.174	0.403	0.193	0.318	0.330
1	0.204	0.389	0.170	0.329	0.321
1	0.175	0.412	0.203	0.349	0.338
1	0.223	0.415	0.200	0.365	0.346
2	0.113	0.142	0.008	0.266	0.214
2	0.134	0.178	-0.009	0.306	0.205
2	0.124	0.145	0.014	0.312	0.198
2	0.118	0.132	0.087	0.300	0.158

2	0.069	0.156	0.045	0.288	0.173
2	0.122	0.122	-0.077	0.291	0.165
3	-0.078	0.102	-0.192	0.308	0.118
3	-0.060	0.156	-0.162	0.291	0.154
3	-0.024	0.122	-0.189	0.295	0.132
3	0.078	0.256	-0.145	0.300	0.106
3	-0.088	0.288	-0.189	0.302	0.171
3	0.046	0.294	-0.174	0.287	0.105
4	0.103	0.303	-0.078	0.288	0.376
4	0.113	0.312	-0.069	0.256	0.380
4	0.098	0.342	-0.080	0.242	0.382
4	0.089	0.308	-0.109	0.276	0.382
4	0.106	0.294	-0.123	0.255	0.370
4	0.117	0.322	-0.080	0.272	0.378
5	0.192	0.377	0.107	0.157	0.284
5	0.182	0.404	0.154	0.149	0.228
5	0.183	0.360	0.113	0.147	0.216
5	0.212	0.384	0.124	0.169	0.272
5	0.223	0.392	0.168	0.140	0.282
5	0.199	0.332	0.182	0.152	0.265

1. Bovine serum albumin 2. Lysozyme 3. Cytochrome c 4. Hemoglobin 5. Myoglobin

Table S7. Data matrix of absorbance response patterns against tissue lysates of normal, Ade, and SCC from 100 ng sensing experiments.

Tissue	NP1	NP2	NP3	NP4	NP5	NP6
1	0.256	0.476	0	0.148	-0.112	-0.342
1	0.330	0.432	-0.018	0.158	-0.045	-0.660
1	0.264	0.456	-0.008	0.088	-0.067	-0.445
1	0.322	0.428	-0.014	0.044	-0.065	-0.830
1	0.292	0.454	-0.002	0.198	-0.104	-0.680
1	0.356	0.448	-0.016	0.132	-0.103	-0.580
1	0.272	0.466	-0.028	0.146	-0.087	-0.820
2	0.472	0.231	-0.077	-0.062	0.011	-0.028
2	0.456	0.203	-0.034	-0.032	0.134	-0.045
2	0.372	0.254	-0.016	0.022	0.117	-0.016
2	0.417	0.238	-0.034	0.101	0.123	-0.013
2	0.443	0.212	-0.030	0.100	0.028	-0.045
2	0.412	0.230	-0.065	0.056	0.009	-0.116
2	0.444	0.259	-0.005	-0.023	0.042	-0.045
3	0.514	0.440	0.222	0.422	0.150	0.014
3	0.534	0.430	0.178	0.406	0.152	0.016
3	0.448	0.416	0.164	0.340	0.100	-0.036
3	0.492	0.452	0.098	0.224	0.038	-0.098
3	0.514	0.486	0.138	0.196	0.034	-0.102
3	0.544	0.454	0.100	0.224	0.180	0.044
3	0.508	0.452	0.150	0.188	0.106	-0.030

1. Adenocarcinoma (Ade) 2. Squamous-cell carcinoma (SCC) 3. Normal

Table S8. Data matrix of absorbance response patterns against different degree of differentiation of Ade and SCC from 100 ng sensing experiments.

Tissue	NP1	NP2	NP3	NP4	NP5	NP6
1	0.256	0.476	0	0.148	-0.112	-0.342
1	0.330	0.432	-0.018	0.158	-0.045	-0.662
1	0.264	0.456	-0.008	0.088	-0.067	-0.447

1	0.322	0.428	-0.014	0.044	-0.065	-0.834
1	0.292	0.454	-0.002	0.198	-0.104	-0.680
1	0.356	0.448	-0.016	0.132	-0.103	-0.582
1	0.272	0.466	-0.028	0.146	-0.087	-0.820
2	0.398	0.464	0.108	0.218	0.012	0.024
2	0.392	0.434	0.148	0.178	0.004	0.052
2	0.374	0.476	0.112	0.202	0.020	0.003
2	0.332	0.478	0.132	0.262	0.028	0.340
2	0.326	0.496	0.140	0.250	0.015	0.002
2	0.338	0.424	0.118	0.148	0.034	0.670
2	0.330	0.444	0.124	0.226	0.005	0.660
3	0.546	0.452	0.134	0.212	0.045	0.002
3	0.574	0.462	0.122	0.216	-0.056	-0.002
3	0.566	0.472	0.138	0.220	0.069	-0.056
3	0.596	0.440	0.104	0.218	-0.060	-0.060
3	0.578	0.480	0.076	0.188	-0.083	-0.043
3	0.558	0.466	0.116	0.246	0.074	-0.028
3	0.516	0.442	0.066	0.212	0.058	-0.025
4	0.472	0.231	-0.077	-0.062	0.011	-0.028
4	0.456	0.203	-0.034	-0.032	0.134	-0.045
4	0.372	0.254	-0.016	0.022	0.117	-0.016
4	0.417	0.238	-0.034	0.101	0.123	-0.013
4	0.443	0.212	-0.030	0.100	0.028	-0.045
4	0.412	0.230	-0.065	0.056	0.009	-0.116
4	0.444	0.259	-0.005	-0.023	0.042	-0.045
5	0.538	0.420	0.168	-0.106	0.118	-0.018
5	0.534	0.402	0.148	-0.046	0.098	-0.068
5	0.522	0.416	0.152	-0.076	0.104	-0.082
5	0.516	0.420	0.164	-0.102	0.110	-0.126
5	0.492	0.390	0.192	-0.124	0.090	-0.046
5	0.490	0.436	0.218	-0.088	0.100	-0.036
5	0.528	0.428	0.174	-0.012	0.152	-0.016
6	0.476	0.296	0.132	-0.004	0.080	-0.034
6	0.424	0.347	0.115	0.027	0.112	-0.036
6	0.399	0.381	0.167	-0.066	0.112	-0.049
6	0.408	0.392	0.119	-0.036	0.046	-0.104
6	0.411	0.331	0.147	0.009	0.132	-0.068
6	0.453	0.356	0.123	-0.046	0.140	-0.042
6	0.442	0.368	0.144	0.013	0.078	-0.060

1. Adenocarcinoma, G1 2. Adenocarcinoma, G2 3. Adenocarcinoma, G3 4. Squamous-cell carcinoma, G1 5. Squamous-cell carcinoma, G2 6. Squamous-cell carcinoma, G3

Table S9. Detection and identification different tissue types or grades of 20 unknown patients, the concentrations of prepared unknown tissue lysates are 100 ng, which is the same with that of in training matrix (table S6, 7).

Tissue	NP1	NP2	NP3	NP4	NP5	NP6	Identification	Verification
1	0.432	0.214	-0.082	0.089	0.125	-0.113	4	4
2	0.500	0.264	-0.003	0.111	0.027	-0.008	4	4
3	0.524	0.476	0.144	0.232	0.055	0.027	3	3
4	0.319	0.456	0.125	0.186	0.020	0.021	2	2
5	0.528	0.424	0.146	-0.138	0.123	-0.093	5	5
6	0.286	0.429	-0.004	0.200	-0.072	-0. 360	1	1
7	0.342	0.473	0.018	0.104	-0.116	-0.715	1	1
8	0.367	0.488	-0.047	0.012	-0.023	-0.285	1	1
9	0.467	0.392	0.169	0.021	0.104	-0.053	6	5 fail

10	0.455	0.372	0.153	0.104	0.117	-0.087	6	6
11	0.396	0.356	0.155	-0.055	0.087	-0.062	6	6
12	0.612	0.484	0.147	0.208	-0.093	-0.034	3	3
13	0.355	0.446	0.106	0.246	0.102	0.001	2	2
14	0.354	0.450	0.122	0.250	0.047	0.068	2	2
15	0.380	0.198	-0.100	0.065	-0.008	-0.076	4	4
16	0.406	0.205	-0.078	0.101	0.122	-0.014	4	4
17	0.483	0.364	0.103	-0.099	0.115	-0.055	4	5 fail
18	0.333	0.406	0.010	0.182	-0.114	-0.457	1	1
19	0.326	0.420	0.117	0.169	0.045	0.127	2	2
20	0.476	0.386	0.115	-0.056	0.085	-0.068	6	6
21	0.515	0.434	0.148	0.219	-0.093	-0.064	3	3
22	0.545	0.432	0.139	-0.121	0.136	-0.122	5	5
23	0.267	0.408	0.025	0.133	-0.083	-0.758	1	1
24	0.490	0.388	0.220	-0.123	0.075	-0.104	5	5
25	0.452	0.301	0.125	0.156	0.047	-0.039	6	6
26	0.409	0.485	0.127	0.176	-0.022	-0.104	2	2
27	0.383	0.164	-0.133	0.099	-0.007	-0.016	4	4
28	0.453	0.331	0.125	0.158	0.143	-0.048	6	6
29	0.562	0.454	0.046	0.208	-0.093	-0.034	3	3
30	0.482	0.410	0.065	0.177	0.065	0.008	2	3 fail

1. Adenocarcinoma, G1 2. Adenocarcinoma, G2 3. Adenocarcinoma, G3 4. Squamous-cell carcinoma, G1 5. Squamous-cell carcinoma, G2 6. Squamous-cell carcinoma, G3

Table S10. The Jackknifed classification matrix showing the contribution of each nanoparticle in the differentiation against different subtypes and grades of NSCLC patients.

NP1	NP2	NP3	NP4	NP5	NP6	%
						50
						48
						60
						43
						33
						40
						100

References

1. Chou SS, De M, Luo J, Rotello VM, Huang J, Dravid VP. Nanoscale Graphene Oxide (nGO) as Artificial Receptors: Implications for Biomolecular Interactions and Sensing. J Am Chem Soc. 2012; 134: 16725-16733.
2. Pei H, Li J, Lv M, Wang J, Gao J, Lu J, Li Y, Huang Q, Hu J, Fan C. A graphene-based sensor array for high-precision and adaptive target identification with ensemble aptamers. J Am Chem Soc. 2012; 134: 13843-13849.
3. Zhu X, Zhang H, Feng C, Ye Z, Li G. A dual-colorimetric signal strategy for DNA detection based on graphene and DNAzyme. RSC Adv. 2014; 4: 2421-2426.