Coding Recognition of the Dose-Effect Interdependence of Small Biomolecules Encrypted on Paired Chromatographic-Based Microassay Arrays

Yifeng Deng*, Zhenpeng Lin, Yuan Cheng

Guangdong Key Laboratory for Research & Development of Natural Drugs, Guangdong Medical University, 524023 Zhanjiang, China

*Corresponding author: dengyf@gdmu.edu.cn

Highlight

Rapid self-recognition of small biomolecules in diverse samples without pre-isolation

Matching of HPTLC-based molecular imprinting and bioautography on microassay arrays

Microarray–based differential expression of substance attributes instead of spot scan

Array auto-collimation of multi-attributes derived from the same 2D-HPTLC component

An array framework for combining phenotype-based and target-based assays with TLC-MS

ABSTRACT

The discovery of small biomolecules suffers from the lack of a comprehensive framework to specifically correlate bioactivities and the contributing small molecules in complex samples with molecular and bioactivity diversity. Here, by mapping a sample's 2D-HPTLC fingerprint to microplates, the paired chromatographic-based microassay arrays are created, which can be used as quasi-chip to characterize multiple attributes of chromatographic components, and as the array differential expression of the bioactivity and molecular attributes of irregular chromatographic spots for interdependent encoding, as well as the automatic-collimated array mosaics of the multi-attributes of each component itself encrypted by its chromatographic fingerprint. Based on this framework, we propose a correlating recognition strategy for small biomolecules with the modulating bioactivities through their self-consistent chromatographic behavior characteristics. In the approach, the small-biomolecule recognition in diverse compounds is transformed into a constraint satisfaction problem, which is addressed through examining the dose-effect interdependence of the homologous 2D code pairs by array matching algorithm, instead of preparing diverse compound monomers of complex test sample for identifying item-by-item. Furtherly, considering the dose-effect interdependent 2D code pairs as links and the digital-specific quasimolecular ions as nodes, an extendable self-consistent framework that correlates mammalian cell phenotypic and target-based bioassay with small biomolecules is established. Therefore, the small molecule contributions and the correlations of bioactivities, as well as their pathway can be comprehensively revealed, so as to improve the reliability and efficiency of screening. This strategy was successfully applied to galangal, and practiced the high-throughput digital preliminary screening of small-biomolecules in a natural product.

KEYWORDS: chromatographic fingerprint, microassay array, LC-MS, small biomolecule, local consistence, correlating recognition

INTRODUCTION

The identification of small biomolecules, including diverse compounds with novel chemical structures, drugs and diagnostic markers, plays an important role in drug discovery and precision medicine[1, 2]. In recent years, advances in omics technologies have begun to enable personalized medicine at an extraordinarily detailed molecular level[3, 4], while research on small biomolecules seems to be much slower. This is partly due to the lack of identification methods that specifically correlate bioactivities and the contributing small molecules in complex samples with molecular and bioactivity diversity.[5, 6].

An earlier well-known identification method for small biomolecules in diverse compounds was bioautography on thin-layer chromatography (TLC)[7, 8]. Recently, Morlock et al. immersed the HPTLC plates with the separated sample extracts in the bacterial culture medium and extracted the components in the bioactive spots through a special TLC-MS interface (Camag) for biotracing[9]. However, due to the interference of the chromatographic matrix on many biological assays, the application of this method has been limited and has not been applied to mammalian cells. Meanwhile, electrophoresis, another type of planar chromatography, has become an important means of omics, because it can be combined with macromolecular immune reactions by transferring gel electrophoresis strips to the electrophilic membranes. Molecular imprinting derived from chromatography, molecular recognition based on binding the template to the functional monomer, is widely used in the fields of chromatographic separation, but it is limited to unidirectional matching of a preset specific target molecule[10].In recent decades, sophisticated artificial molecular diversity libraries and various cutting-edge microchips, such as small molecular microarrays and microfluidic chips[11, 12], have emerged; these tools are being used for biological, chemical or physical assays, with wide applicability, high throughput and easy automation features[13]. However, these assays are highly dependent on molecular diversity libraries and are seldom

used for practical diverse compounds, such as natural products and clinical samples. Because it is difficult to capture the diversity of molecules of real samples, such as natural products and clinical tissues, and these diverse compounds have to be separated to monomers regardless of whether they are bioactive[14]. On the other hand, mass spectrometry imaging (MSI) has emerged as the future frontiers of chemical analysis[15], which combines the information gained from mass spectrometry (MS) with visualization of spatial distributions in thin sample sections. MS has evolved to identify diverse compounds as their respective digital-specific quasimolecular ions. For small molecules in a certain sample, a digital-specific quasimolecular ion is simply used for the identification of a compound. However, these methods have not yet been integrated to reveal the correlation between molecular characteristics and bioactivities. As such, it is difficult to adopt any single bioassay or chemical analysis to reflect the molecular and bioactivity diversities of a real sample. How to "integrate both the phenotypic and target-based approaches to estimate a relevant network from compound to phenotype in screening" is a clear challenge[16, 17].

With the success of today's omics research, specific recognition methods for biomacromolecules in complex samples play an important part. The most important aspect of these methods is the correlation between molecular characteristics and bioactivities, such as the pairing of bases encoded on DNA arrays[18] and antigen-antibody immunoreactions based on the "molecular recognition exhibits molecular complementarity" practice[19]. Recently, the world's first actual photo of the entangled quantum behavior of photon pairs of the same beam separated by a polarizing crystal was captured[20]. The entangled quantum photon behavior can be used as random encrypted keys for remote communication, and studies on macroscopic entanglement have emerged[21, 22]. Realistically, NASA's night-light images of the Earth have been used to map the human socioeconomic behavior for some years. Inspired by these approaches of correlating specific behavioral observation, we realize that the

chromatographic behavior characteristics of a small molecule can also be used to correlate its own multiple attributes, such as molecular properties and modulating bioactivity, and thus propose a correlating recognition strategy for small biomolecules.

Around this research strategy, this article discusses the following issues: characterizing multiple attributes of chromatographic components in irregular chromatographic spots as microarray distribution gradients, transforming the microassay array results into two-dimensional codes and self-recognizing biomolecules with array matching algorithm, etc., and the scientific validity, the error control and the affecting factors of this strategy are also discussed.

EXPERIMENTAL METHODS

Materials and Reagents. Galangal (Alpinia officinarum Hance) was Collected in Longtang Town, Guangdong Province, China. Galangin and Diphenylheptane A was purchased from NICPBP (Lot No. 111699-2006021 and 11757-200601). 5-Fluorouracil was purchased from Shanghai EKEAR (Lot No. 150729). Gold Nanoparticles (GNPs) prepared with reference to the literature.30 AlamarBlue™ Cell Viability Reagent was purchased from Invitrogen (Cat#DAL1025). Human A549 cells and Human HepG2 cells was provided by Clinical Research Center of Guangdong Medical University. Oligonucleotides (G-DNA): 5'-TTA GGG TTA GGG TTA GGG TTA GGG-3' was synthesized by Sangon Biotech (Shanghai Bioengineering Co., Ltd., order No. 300064096). The reagents used in the experiment were all analytical reagents.

Establishment of a 2D-HPTLC fingerprint developed in the microarray format. Dry powder (250 μm) of *Alpinia officinarum* Hance (collected in Longtang Town, Guangdong Province, China) was extracted by conventional ultrasonication with petroleum ether, ethanol, tetrahydrofuran, acetone, dichloromethane, ethyl acetate, benzene and water (each

representing one of the 8 kinds of solvents classified by Snyder[23]). The extracts were combined to obtain a representative extract of galangal.

The sampling amounts were chosen by conventional pre-bioassays of the crude sample extract. Generally, the spotting volume of the crude extract on the HPTLC plate is controlled to be no more than 10 μL. The selected amount of the extract was spotted to achieve a 2.5 mm diameter on an HPTLC plate (Merck, HPTLC silica gel 60, aluminum sheets, GF254). HPTLC was extended to 7.2 cm × 7.2 cm according to the 384-well microplate array format (4.5 mm×4.5 mm matrix; the opening of the well of a 384-square-well microplate is approximately the size of a typical HPTLC spot), and the chromatographic two-dimensional mobile phase was used to establish a 2D-HPTLC fingerprint. The optimized two-dimensional mobile phase[24] used for 2D-HPTLC consisted of trichloromethane/MeOH/petroleum ether at a ratio of 7.76/0.24/2 (v/v/v) and ethyl acetate/petroleum ether/acetic acid at a ratio of 3/7/0.2 (v/v/v).

Chromatographic-based microassay array preparation. According to the microarray format of the high-throughput microplate, the silica gel layer of the 2D-HPTLC fingerprint is defined and differentiated by the microarray into a square chip array[25], and the components on the square chip array are mapped into a microplate as a chromatographic-based microassay array with chromatographic matrix removed. Specifically, the silica gel layer was split and stripped from a thin aluminum sheet with a metal grid interface having a 384-well microplate array format (3D printing, original equipment manufacturer (OEM) according to the drawing, Supporting Information S1) under a parallel pressure of 8×10^3 kPa (Specac Atlas® Manual Hydraulic Press-15T) to fabricate a flat square chip array (Figure 1, Supporting Information S2). The chip array was positioned into a corresponding 384-square-well filter plate (Pall 5072), and the components were eluted with methanol into another 384-well microplate. Note that the G-quadruplex ligand bioassay should be pre-washed three

times with 15 μL of water per well to remove possible salts. The microplate containing the eluent was evaporated to dryness in a vacuum centrifuge concentrator to prepare a stock chromatographic-based microassay array (Supporting Information S3). A schematic diagram of the preparation and workflow of the chromatographic-based microassay arrays is presented (Figure S3). The paired chromatographic-based microassay array with the consistent chromatographic distribution is prepared by transferring a portion of the sample from the stock chromatographic-based microassay arrays to the corresponding array units of another 384-well microplate.

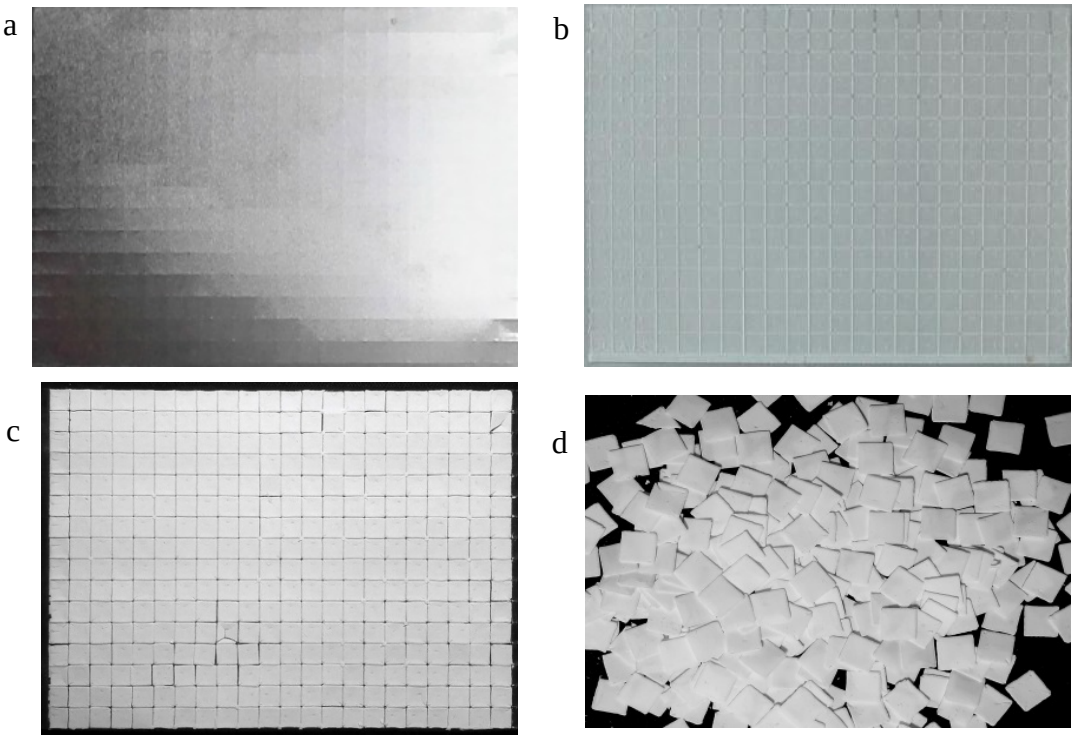


Figure 1. Array division and stripping effect of a HPTLC fingerprint. (a) Front and (b) back sides of the aluminum HPTLC sheet that had been pressed by the metal grid interface, (c) square silicon layer array stripped off and adhered to a PVDF film (painted black), and (d) scattered square silica gel layer chips.

Cell viability assays. The A549 and HepG2 cell lines were obtained from the Affiliated Hospital of Guangdong Medical University. With reference to Invitrogen's AlamarBlue® Cell Viability Assay Protocol and related literature methods[26, 27], three groups of experiments were set up as follows: blank group, positive control group and sample group. Then, 0.1% DMSO and 1% glycerol were used as the reagent blank group, while 5-fluorouracil (Shanghai EKEAR, Lot No. 150729) at a concentration of 0.077 mmol•L^{-1} was used as a positive control. The amount that adequately reflected the bioactivity of the chromatographic-based microassay array sample was chosen through the cell viability pre-bioassay as the test dose for subsequent bioassays[24].

The cells were cultured in high-glucose DMEM with 8% fetal bovine serum at a cell density of 1,000 cells per 40 μL in each well of the 384-well microplate. The cells were incubated at 37°C under 5% CO_2. At 12 hours, 20 μL of the cell culture medium containing the optimal amount of test sample was added to the corresponding well of the microplate, and the cells were further incubated for 36 hours. Then, each group was centrifuged (12 ×g, 15 s) to remove the drug-containing cell culture medium and the cells were washed twice with PBS and discarded after centrifugation. Fifty microliters of cell culture medium and 5 μL of AlamarBlue (Invitrogen, Cat#DAL1025) were added per well to each group, and the samples were degassed with a vacuum centrifugal concentrator (269 ×g, room temperature, -80 KPa, 1 min). The absorbance of the microplate was measured at 570 nm and 600 nm (Bio-Tek Epoch one). Then, the cells were further incubated at 37°C under 5% CO_2 for 12 hours. The absorbance was measured again using the same method. The difference between the absorbance of the cells incubated 12 hours after the addition of AlamarBlue and the absorbance at the time of initial addition was used as the absorbance value, and the cell survival rate was calculated according to Invitrogen's protocol for the AlamarBlue® Cell Viability Assay.

G-quadruplex ligand bioassay. Referring to the literature methods and with slight changes[28, 29], we used gold nanoparticles (GNPs) as a colorimetric probe and screened G-quadruplex ligands by colorimetry based on the color differences between single-stranded DNA-GNPs and G-quadruplex-GNPs. The amount that adequately reflected the bioactivity of the chromatographic-based microassay array sample was chosen through the G-quadruplex ligand pre-bioassay as the test dose for subsequent bioassays. The initial total volume is 35μL per well, which contains 5μL methanol. As previously described in the literature[30], the GNPs were prepared by reducing chloroauric acid with sodium citrate. The diameter was calculated to be 12 nm, and the concentration was 9.45 nM. The GNPs and 1.0 μM GDNA (5'-TTA GGG TTA GGG TTA GGG TTA GGG-3', synthesized by Shanghai Bioengineering Co., Ltd., order no. 300064096) were mixed at a 1:150 molar ratio (optimized through an anti-NaCl-induced aggregation experiment performed by microplate titration, in which the range of the molar ratio of GDNA to GNPs was from 50-250) and incubated at room temperature (25°C) for 16 hours to prepare the GNP-GDNA probe. Then, 60 μL of the GNP-GDNA probe per well was added to the chromatographic-based microassay array. After 3 hours of incubation at room temperature (25°C), NaCl solution was added to each well in the microplate containing the reaction solution to achieve a concentration of 0.106 M, which was followed by further incubation for 30 min at room temperature (25°C). After the solution was added, the microplate was treated with a vacuum centrifugal concentrator (269 × g, at room temperature, -80 kPa, 1 min) to remove bubbles. The absorbance spectra (400~850 nm) of each well of the microplate were collected (Bio-Tek Epoch one). The color change was observed, and the absorbance ratio calculated at 670 nm and 520 nm was used to evaluate the ligand activity of the components in each array unit of the chromatographic-based microassay array.

LC–ESI-MS analysis. According to the bioactivity heat map, a portion of the sample corresponding to the significant bioactive array units is transferred from the stock chromatographic-based microassay array to the corresponding array units of another 384-well microplate for LC-ESI-MS analysis, which was conducted with an Agilent 6430 triple quadrupole mass spectrometer coupled with an electrospray ionization (ESI) source. The operating conditions were as follows: gas flow rate of 12 L•min^{-1}, gas temperature of 350°C; sheath gas flow rate of 12 L•min-1, sheath gas temperature of 350°C and capillary voltage 3.6 kV. The mass analyses were performed with an ESI source in negative ionization mode; the m/z scan range was set from 65 to 750. The chromatographic system was composed of an Agilent 1200 series HPLC. The eluent was 75% methanol-H_2O introduced at a flow rate of 0.2 mL•min^{-1}. The injection volume was 5 μL in direct injection mode. Agilent Mass Hunter software was used for data analysis.

RESULTS AND DISCUSSIONS

Protocol design for the correlation between molecules and bioactivities. The paired chromatographic-based microassay arrays are created from an array-differential sample's 2D-HPTLC fingerprint with the chromatographic matrix removed and the consistent chromatographic distribution (Supporting Information). This paired chromatographic-based microassay arrays can be used as quasi-chip to characterize multiple attributes of chromatographic components, and as the array differential expression of the bioactivity and molecular attributes of irregular chromatographic spots for coding interactions, as well as the automatic-collimated array mosaics of the multi-attributes of each component itself encrypted by its chromatographic fingerprint.[25, 31]. A bioassay, i.e., a cell-based phenotypic assay or ligand binding assay, is performed using one of the paired chromatographic-based microassay arrays, and biotracing is followed correspondingly by direct ESI-MS analysis on the other one. Thus, local self-consistency of the array distribution of bioactivity strength vs. the quasimolecular ion intensity of the modulating compounds on the paired chromatographic-based microassay arrays is established. A schematic diagram of the workflow used for the strategy is shown (Figure 2).

In general, 2D-HPTLC may not be able to achieve complete separation of all the chromatographic spots of diverse compounds, and due to the shape, size and array distribution of HPTLC spots are irregular, each chromatographic spot does not always fall into a single microarray unit, and several chemical components may coexist in the same microarray unit. However, on the paired chromatographic-based microassay arrays, the chromatographic spots, even if they are partially overlapped, are differentiated by the microarray with vector feature. By virtue of the high resolution and high sensitivity of LC-ESI-MS, each digital specific quasi-molecular ion of coexisting small molecule compounds is characterized as its intensity array distribution gradient, and the trend difference of the array

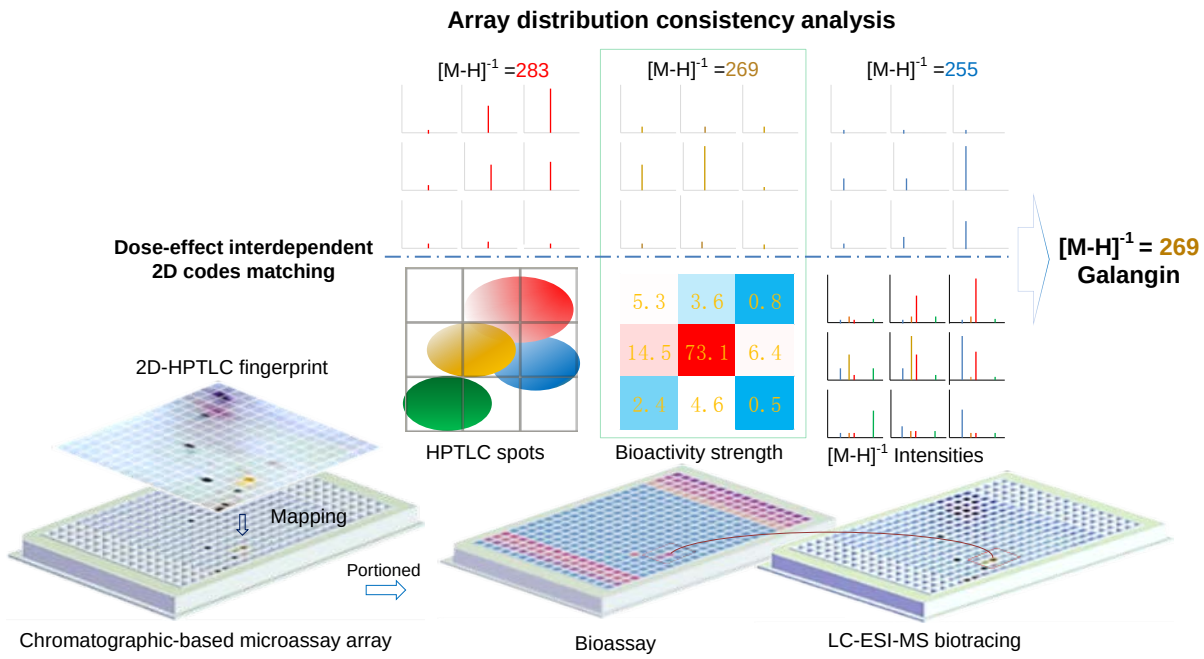


Figure 2. Schematic diagram of the workflow for establishing the array local self-consistency

distribution gradients between the bioactivity strength vs. the digital-specific quasimolecular ion intensities of several coexisting components can be clearly distinguished in the corresponding regions (Figure 2). The bioactive compound must be present in those array units that exhibit bioactivity. When multiple components coexist there, it is important how to identify the molecule attributed to bioactivity from them. This experimental design provides a satisfactory solution.

Dose-effect interdependence is a basic pharmacological principle. Under this experimental design, the array distribution gradient between the bioactivity strength and the specific quasimolecular ion intensities of the respective modulating compound are local self-consistent, the attributes derived from the same compound will be auto-collimated, regardless of the irregular shape, size and array distribution of its chromatographic spot. Therefore, the experimental data obtained from the paired chromatographic-based microassay arrays are correspondingly extracted as homologous 2D code pairs, and are substituted into the array matching algorithm to determine whether there is dose-effect interdependence, and to identify

the digital-specific molecular characteristics of the compound attributed to the bioactivity among the coexisting components in the corresponding array region[32, 33].

As a result, the target deconvolution, an important aspect of current drug discovery[34], is simplified as an array matching algorithm based on the principle of dose-effect interdependence, just like addressing constraint satisfaction problems (CSPs).

Resolution Analysis and Error Control. In order to reflect the array distribution gradient of the chromatographic spots in the 2D-HPTLC fingerprint by a chromatographic-based microassay array with a high degree of sharpness, the array unit is preferably designed to resemble chromatographic spots in size to avoid more than one separated chromatographic spots falling into the same array unit as a whole or the same chromatographic spot being split into similar array units in terms of chemical composition. It is feasible and practicable to characterize 2D-HPTLC fingerprint spots (typically approximately 2.5 mm to 4.5 mm in diameter) with commercial 384-square well high throughput microplates (with an array format of 4.5 mm×4.5 mm/well, 16 rows×24 columns), such as human cell-based phenotypic assays (typically 40 μL per well, with wells containing 1000 cells), target-based assays and LC-ESI-MS biotracing. The array distribution of the chromatographic spots (of the compounds developed with the sulfuric acid-vanillin reagent) in a 384-well array grid is shown (Figure 3).

Cell viability assays are the bottleneck of the resolution, which needs a certain space for cell growth, and the components of chromatographic spots should reach the drug dose range. As the RSD of the Alamar Blue cell viability assay used in this experiment is about 8%, this requires that the cell viability of the detected array units is more than 24% (3 times the RSD). Since the content of chromatographic components in the array unit may fluctuate between 1 ~ 1/4 (in the worst case, one spot may be evenly distributed in four array units), so the proportion of non-overlapping parts of chromatographic points must be greater than about

25%. This means that the resolution ratio of chromatographic spots should be greater than 0.5 (As the resolution ratio reaches 1.5, the adjacent two spots reach baseline separation). In terms of planar chromatography technology, such resolution requirements are perfectly achievable.

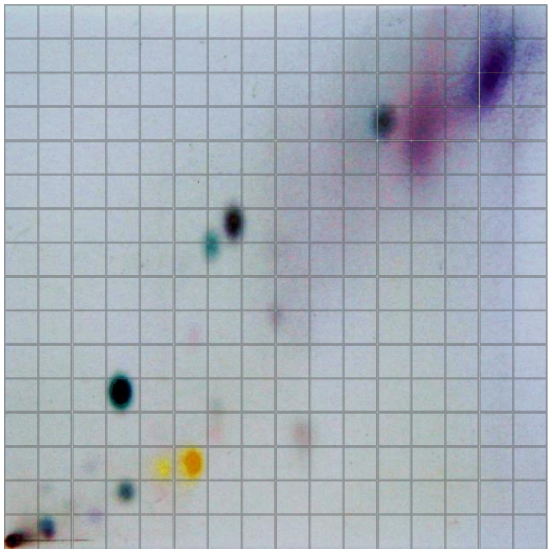


Figure 3. Schematic diagram of the array distribution of the 2D-HPTLC chromatogram of galangal extract

In fact, the worst case mentioned can be avoided. Since on the HPTLC plate (silica gel 60, GF254) used in this experiment, most of the chromatographic spots of biomolecules can be detected non-destructively under UV light, and the bioactive regions can be known during pre-bioassays for sample loading amount, so the bioactive regions can be pre-located in as few array units as possible by adjusting the relative position between the array grid and 2D-HPTLC fingerprint.

Generation of 2D codes from obtained experiment results. This method was applied to galangal (*Alpinia officinarum* Hance, a famous traditional Chinese medicine) as a paradigm. The anti-cancer activity of galangal or its mechanism of action has been reported one after another[35, 36]. Focused on the anticancer effects of galangal, the bioassays of this experiment include the A549 and HepG2 cell viability assays and G-quadruplex ligand

bioassay, and were performed on the chromatographic-based microassay arrays. The obtained bioactivity strengths are expressed in the array distribution and are then converted into two-dimensional codes by Excel (2010 version). The bioassay data and the generated two-dimensional codes are shown (**a**, **b** and **c** in Figure 4).

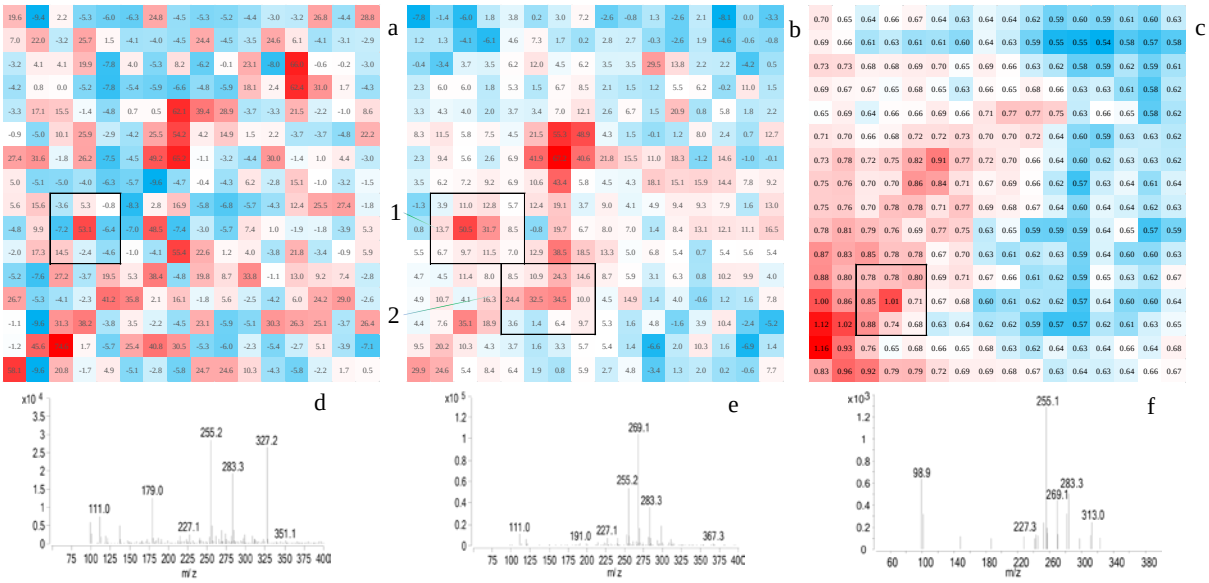


Figure 4. Array distribution of bioactivity strengths (2D code) generated from the results of the bioassay and the MS spectrum of the central array unit corresponding to the significant bioactive regions marked with a wireframe. (**a**) 2D codes generated from the viability data of the A549 cell line; (**b**) 2D codes generated from the HepG2 cell line; and (**c**) 2D codes generated from the G-quadruplex ligand bioassay. The MS spectra corresponding to the center array units of the significant bioactive regions marked with a wireframe in **a**, **b** and **c** are shown in **d**, **e**, **f**, respectively.

Correspondingly, the compositions of the array units with significant bioactivity are further analyzed *in situ* by LC-MS and are highly resolved into their digital-specific quasimolecular ion intensities in the array distribution, and the MS spectra of the central array unit corresponding to the significant bioactive regions are obtained (**d**, **e**, and **f** in Figure 3, marked with wireframe. The chromatographic-based microassay arrays for the bioassay

and LC-MS bioactivity tracing should be sampled from the same chromatographic-based microassay array stock sample as the homologous test samples). The main coexisting quasimolecular ions are determined based on the mass spectrum corresponding to the central array unit of the significant bioactive region. The array distribution data of the quasimolecular ion intensities of the main coexisting quasimolecular ions are extracted from the CSV file of the mass spectra of the array units corresponding to the significant bioactive region, and the array distribution of the quasimolecular ion intensity of each coexisting component is obtained. Then, the array distribution of the specific quasimolecular ion intensities is converted into the corresponding two-dimensional code by Excel (2010 version).

In this experiment, the analytical method of combining 2D-HPTLC with LC-ESI-MS is a powerful means of identifying small molecular compounds. 2D-HPTLC has strong chromatographic distinguishing power and can even be used to identify isomers[37, 38]. LC-ESI-MS is the preferred analytical method for complex samples and serves as a benchmark for other mass spectrometry methods due to its representativeness, reproducibility and quantitative comparability[39]. Furthermore, electrospray ionization (ESI), a "soft" ionization technique that rarely damages the structure of compounds during ionization, can be used to directly identify the small molecular compound[15]. In addition, bioactive-directed LC-ESI-MS tracing on paired chromatographic-based microassay arrays is very effective because it is focused on several bioactive array units rather than on the entire microarray, the workload is greatly reduced.

Dose-effect interdependence assessment. Dose-effect interdependence assessment of the array distribution gradients between the bioactivity strength and the specific quasimolecular ion intensities is performed with the CORREL (Array1, Array2) function in Excel (2010 version). The bioactive strength data array from the significant bioactive region and the adjacent region is selected as the variable Array1, while the quasimolecular ion

intensity data arrays from the coexisting quasimolecular ions in the corresponding region are selected as the variable Array2. These experimental data are substituted into the CORREL (Array1, Array2) function as two-dimensional code pairs to assess the dose-effect interdependence. This method can be regarded as a prototype of the algorithm for this constraint satisfaction problem (CSP), which is used to evaluate the correlation between molecules and bioactivities in diverse compounds.

The data of the quasimolecular ion intensities, bioactivity strengths and the correlation analysis results obtained for the coexisting components in the four bioactive regions (**a**, **b**1, **b**2 and **c** in Figure 4, marked with a wireframe) are shown (Figure 5). We can visually observe which coexisting compound has an array distribution gradient of specific quasi-molecular ion intensity that tends to be consistent with the array distribution gradient of bioactivity intensity in the corresponding array region (Figure 5). Using array correlation algorithm, such as Excel CORREL function, to calculate correlation coefficient can make the array difference objective and quantitatively comparable, so as to facilitate computer-aided analysis. The absolute value of the correlation coefficient is used to rank the compounds corresponding to the coexisting quasimolecular ions for bioactivity screening. The larger the value, the more significant the difference and the higher the credibility. More obviously, if the array unit with the maximum quasimolecular intensity deviated significantly from the array unit with the maximum bioactive strength in the corresponding region, the corresponding compound would not be considered the bioactive compound.

Intensity array in Figure 4 a bioactive region

Cell assay onA549 cell line（%）			[M-H]$^-$ intensity Array ×10^{-5} 327			255			283		
-3.6	5.3	-0.8	0.34	0.56	0.40	2.30	2.55	2.82	1.51	1.63	1.74
-7.2	53.1	-6.4	0.83	2.76	1.20	2.89	2.80	3.92	1.78	1.80	2.42
14.5	-2.4	-4.6	0.47	0.66	0.33	3.00	3.45	3.50	1.90	2.05	2.38
γ			0.827			-0.256			-0.224		

Intensity array in Figure 4 b1 bioactive region

Cell assay on HepG2 cell line（%）				[M-H]$^-$ intensity Array ×10^{-5} 327				255				283			
3.9	11.0	12.8	5.7	0.18	0.85	1.03	0.08	3.32	2.98	3.40	4.57	1.77	1.60	1.69	2.00
13.7	50.5	31.7	8.5	0.22	1.29	2.69	0.08	2.82	2.75	2.86	4.10	1.45	1.52	1.58	1.88
6.7	9.7	11.5	7.0	0.11	0.19	0.27	0.07	2.57	3.10	3.82	10.1	1.52	1.64	1.89	4.27
γ				0.712				-0.288				-0.268			

Intensity array in Figure 4 b2 bioactive region

Cell assay on HepG2 cell line（%）				[M-H]$^-$ intensity Array × 10^{-5} 269				283				255			
8.5	10.9	24.3	14.6	1.69	2.20	1.52	1.13	2.02	3.08	2.34	1.96	4.55	3.52	5.30	4.35
24.4	32.5	34.5	10.0	1.25	8.19	3.02	1.26	1.67	2.93	2.75	2.50	3.27	3.99	5.88	6.37
3.6	1.4	6.4	9.7	1.31	2.35	1.44	0.64	1.43	2.03	2.29	1.46	2.68	3.75	4.98	3.05
γ				0.560				0.454				0.305			

Intensity array in Figure 4 c bioactive region

G-quadruplex ligand assay			[M-H]$^-$ intensity Array ×10^{-3} 269			255			283		
0.88	0.86	0.78	0.34	0.40	0.48	1.69	1.56	1.74	1.12	0.92	1.06
0.74	1.01	0.78	0.62	1.12	0.50	1.09	1.90	2.00	0.76	0.94	1.14
0.68	0.72	0.80	0.20	0.27	0.52	0.99	1.51	2.69	0.62	0.83	1.91
γ			0.758			0.411			0.209		

Figure 5. Dose-effect interdependence assessment of the array distribution gradients between bioactive strength and quasimolecular ion peak intensities. γ-Correlation coefficient of the array distribution of bioactive strength with the array distribution of quasi-molecular ion peak intensity obtained from the Excel CORREL function. The background color of the cells in the table is generated according to the cell values. The larger the value is, the darker the blue color; the smaller the value is, the darker the red color.

Bioactive candidates and bioactivity evaluation. Based on the calculated correlation coefficients, compounds corresponding to 327 $[M-H]^-$ and 269 $[M-H]^-$ were identified as bioactive candidates by the cell viability assay, and 269 $[M-H]^-$ was identified as a bioactive candidate by the G-quadruplex ligand bioassay. The bioactivity strength of the same quasimolecular ion in the array unit within the same bioactive array region was normalized to the contribution of this quasimolecular ion and incorporated into the bioactivity screening consideration. Since the small molecules causing bioactivity are identified from the original sample rather than from the reaction products, the abundant database of natural products can provide support. A literature search and a compound database search for bioactive candidates were performed. Referring to the literature[40], 327 $[M-H]^-$ and 269 $[M-H]^-$ in the galangal extract were speculated to belong to diphenylheptane A and galangin (CAS No. 68622-73-1 and 548-83-4), respectively.

The cell bioactivities of the two candidate compounds were evaluated by cell viability assays with 5-fluorouracil as a positive control. The cell growth inhibition curve data were analyzed with SPSS 18.0 to calculate the IC_{50} values. The IC_{50} values of diphenylheptane A (NICPBP, Lot No. 111757-200601), galangin (NICPBP, Lot No. 111699-200602) and 5-fluorouracil for the A549 cell line are 0.247, 0.089, and 0.023 mmol•L-1, respectively, and those for the HepG2 cell line are 0.259, 0.085, and 0.092 mmol•L-1, respectively. There was no significant difference between the bioactivity of galangin-induced G-quadruplex DNA formation (0.659±0.038, n=8) and that of aloe emodin (NICPBP, Lot No. 0795-9702), which was used as a positive control (0.686±0.026, n=8), at concentrations of approximately 0.5 μM.

The scientificity and effectiveness of the strategy In this strategy, a pair of chromatographic-based microassay arrays are created to characterize and assemble the modulating bioactivity and molecular characteristics of the compounds separated on 2D-HPTLC into self-consistent array distribution gradient, the contribution of small molecules to

bioactivity was assessed according to the basic pharmacological principle of dose-effect interdependence, and all the experiments employed the conventional proven pharmaceutical research methods and specifications. Ultimately, the screened candidates were determined by bioactivity assessment compared with bioactive control. Therefore, the results obtained by the strategy are scientific and valid.

Due to the combined multidisciplinary approaches, this strategy has achieved a lot of research innovations. For the first time, the total compounds separated on the HPTLC fingerprint were subjected to mammalian cell-based phenotypic assay without interference from the chromatographic matrix; the observation indicators are converted from conventional chromatographic spots to array distribution gradients of essential substance attributes, e.g. apoptotic cell morphology and digital-specific quasimolecular ion, and the resolution is greatly improved; The compounds' attributes are characterized in intensity array distribution gradient and digitized into homologous 2D codes, so the small biomolecules can be recognized through array matching algorithm with the assistance of computer. Therefore, high-throughput digital preliminary screening of small biomolecular in diverse compounds can be realized without having to separate the test sample into monomers, the screening efficiency is significantly improved, and the workload is greatly reduced.

Another notable aspect, this experiment shown that galangin in *Alpinia officinarum* Hance has bioactivities for inducing G-quadruplex DNA formation and inhibiting cancer cell proliferation, while diphenylheptane A also inhibits cancer bioactivity. These results suggest that galangin induces the formation of G-quadruplex DNA and is involved in the inhibition of cancer cell proliferation, which may be one of the pathways of galangin's anticancer activity. Although diphenylheptane A also has anticancer activity, its mechanism of action is different from that of galangin.

It is important to "integrate both the phenotypic and target-based approaches to estimate a relevant network from compound to phenotype in screening"[16, 17]. In this experiment, human cell phenotypic (two cell lines) and ligand based bioassay were coupled with LC-ESI-MS biotracing on the paired chromatographic-based microassay arrays. By considering the digital-specific quasimolecular ions as nodes and dose-effect interdependent code pairs as links, the cell phenotypic and target-based bioassays can be integrated with LC-ESI-MS biotracing to establish an extendable local self-consistent framework. Thus, the small molecular contributions and correlations of bioactivities, as well as their pathway can be comprehensively revealed. This is helpful to eliminate the false positive in screening and significantly improve the reliability of small biomolecules recognition.

Factors affecting self-consistent analysis. In this experiment, the correlation coefficient is not as large as the normal standard calibration curve, and some microarray regions that appear to exhibit bioactivity fail to resolve molecules attributed to bioactivity (Figure 4). There may be particular reasons, such as the high content, the linear range of the signal response, and the memory effect of the thin-layer silica gel. The components of natural products are diverse and vary greatly in content. The concentration levels of TLC spots prepared using a certain sampling amount are not all suitable for bioassays, and the ultrahigh concentration levels in some microarray units may lead to nonspecific bioactivity effects. In addition, these phenomena can occur if the local content is too high due to spot overlap. The correlation coefficient here reflects the multifactorial correlation between the response signals of the bioassay system and those of the LC-ESI-MS system. If one of the respective signal responses detected by the bioassay or LC-ESI-MS is not within the linear response range, the correlation may weaken. In addition, due to the memory effect, trace components remaining in the thin-layer silica gel can be strongly eluted by methanol. The interferences from these components will appear as noise detected by LC-MS or bioassays. This type of memory

residue of the complex components is not the same, and interference cannot be completely deduced from the reagent blank. After all, the chemical composition on the array unit of the chromatographic-based microassay array is usually not a single component, and it also involves multi-factor response values. The correlation coefficient mentioned here is different from the correlation coefficient of the standard calibration curve obtained for the standard reference material.

CONCLUSIONS

Through building the paired chromatographic-based microassay arrays, high-throughput micro-separation, *in situ* recording, μTAS, coding interaction and correlating recognition of substances are achieved, and an extendable local self-consistent framework that combining mammalian cell phenotypic and target-based bioassay with HPTLC-MS is established. Therefore, the small molecular contributions and correlations of bioactivities and their pathways can be associated for comprehensive screening. This approach can make use of abundant resources, such as validated drug research methods, drug databases and TLC fingerprint library recorded in pharmacopoeia. The digital acquisition and identification algorithm of chromatographic data is conducive to the computer-aided processing of big data, and has the prospect of artificial intelligence. This research strategy will greatly promote the drug discovery of small biomolecules.

ASSOCIATED CONTENT

Supporting Information

Supporting Information is available. The information includes preparation of the metal array grid interface, preparation of the square chip array of the silica gel layer, and elution of the silica gel chips with a 384-square-well filter plate. A schematic diagram of preparation of chromatographic-based microassay array can be found in the online Supporting Information. Figures S1−S3 are shown.

AUTHOR INFORMATION

Corresponding Author

*Email: dengyf@gdmu.edu.cn

AUTHOR CONTRIBUTIONS

Yifeng Deng conceived the project, designed all the experiments in this study and drafted this manuscript for publication. In addition, he performed all the experiments for the G-quadruplex ligand bioassay and consistency analysis of all the experimental data. Zhenpeng Lin performed the cell viability assays and the LC-MS analysis, processed the related data, prepared some Chromatographic-based microassay arrays, and created some of the CAD drawings. Yuan Cheng carried out the preliminary experiments for this study.

ACKNOWLEDGMENTS

We thank Yu Zhong (Analysis Centre of Guangdong Medical University, China) for technical assistance during LC-MS analysis, Liubo Lan for help with the initial cell viability assay, and Ning Li (Affiliated Hospital of Guangdong Medical University, China) for providing the A549 and HepG2 cells. This study was financially supported by the Bureau of Guangdong Traditional Chinese Medicine, China (No. 20141152 and No. 20181153).

NOTES

The authors declare that they have no competing financial interests.

REFERENCES

(1) Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G. Counting on natural products for drug design. *Nature Chemistry* **2016,** *8*, 531–541.

(2) World Health Organization, WHO traditional medicine strategy: 2014-2023. **2013**.

(3) Karczewski, K. J.; Snyder, M.P. Integrative omics for health and disease. *Nature Reviews Genetics* **2018,** *19*, 299–310.

(4) Weston, A. D.; Hood, L. Systems Biology, Proteomics, and the Future of Health Care: Toward Predictive, Preventative, and Personalized Medicine. *Journal of Proteome Research* **2004,** *3*, 179-196.

(5) Schenone, M.; Dančík, V.; Wagner, B.K. Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology* **2013,** *9*, 232–240.

(6) Swinney, D. C.; Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **2011,** *10*, 507–519.

(7) Helgeland, K.; Jonsen, J.; Laland, S.; Lygren, T.; & Romcke, O. Biological activity of a brownish-yellow pigment produced from vitamin B12 by Aerobacter aerogenes. *Nature* **1963,** *199*, 604-605.

(8) Bjørseth, A.; Eidså, G.; Gether, J. Detection of mutagens in complex samples by the salmonella assay applied directly on thin-layer chromatography plates. *Science* **1982,** *215*, 87-89.

(9) Jamshidi-Aidji, M.; & Morlock, G. E. From Bioprofiling and Characterization to Bioquantification of Natural Antibiotics by Direct Bioautography Linked to High-Resolution Mass Spectrometry: Exemplarily Shown for Salvia miltiorrhiza Root. Analytical Chemistry **2016,** *88*(22), 10979–10986.

(10) Maryam Arabi, Abbas Ostovan, Ahmad Reza Bagheri, Xiaotong Guo, Liyan Wang,Jinhua Li, Xiaoyan Wang, Bowei Li, Lingxin, Chen.Strategies of molecular imprinting-based solid-phase extraction prior to chromatographic analysis. *TrAC Trends in Analytical Chemistry* **2020**, *128*,115923.

(11) Ma, H.; Horiuchi, K.Y. Chemical microarray: a new tool for drug screening and discovery. *Drug discovery today* **2006,** *11*, 661-668.

(12) Rothbauer, M.; Wartmann, D.; Charwat, V.; Ertl, P. Recent advances and future applications of microfluidic live-cell microarrays. *Biotechnology Advances* 2015, *33*, 948–961.

(13) Schneider, G. Automating drug discovery. *Drug Discovery* **2018,** *17*, 97–113.

(14) Janzen, W. P. Screening technologies for small molecule discovery: the state of the art. *Chemistry & biology* **2014,** *9*, 1162-1170.

(15) Buchberger, A. R.; DeLaney, K.; Johnson, J.; & Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. Analytical Chemistry **2017,** 90, 240–265.

(16) Glenn E. Croston. The utility of target-based discovery. *Expert Opinionon Drug Discovery* **2017**,12, 427-429.

(17) Iwata, H.; Kojima, R.; & Okuno, Y. An in Silico Approach for Integrating Phenotypic and Target-Based Approaches in Drug Discovery. *Molecular Informatics* **2020**, 39, eLoc. 1900154.

(18) Hood, L.; Galas, D. The digital code of DNA. Nature **2003,** 421, 444–448.

(19) Gellman, S. H. Introduction: molecular recognition. Chem. Rev. **1997,** 97, 1231–1232.

(20) Moreau, P.-A.; Toninelli, E.; Gregory T.; Aspden, R. S.; Morris, P. A.; & Padgett, M. J. Imaging Bell-type nonlocal behavior. *Science Advances* **2019**, 5(7), eaaw2563.

(21) Zaminpira, S. The relation between the quantum entanglement in theoretical physics as a new insight into the cancer biology. *Bulletin of the American Physical Society*. APS April Meeting **2019**, 64, 3.

(22) McFadden, J.; Al-Khalili, J. The origins of quantum biology. *Proc. R. Soc. A.* **2018**, 474(2220).

(23) Snyder, L. R. Classification of the solvent properties of common liquids. *Journal of Chromatography A* **1974**, *2*, 223-230.

(24) Cheng, Y.; Li, Y.; Li, J.; Deng, Y. A rapid, micro-scale preliminary screening method for active components in Galangal with protective effect against hydrogen peroxide induced cell apoptosis through “thin layer chromatography” and “tetrazolium-based colorimetric assay” array correspondence. *Journal of Chromatography A* **2015**, *1395*, 167-172.

(25) Geng, Y.; Peveler, W. J.; & Rotello, V. Array-based “Chemical Nose” Sensing in Diagnostics and Drug Discovery. *Angewandte Chemie International Edition* **2019**, *58*, 5190-5200.

(26) Shimada, K.; Skouta, R.; Kaplan, A.; Yang, W.S.; Hayano, M.; Dixon, S.J.; … Stockwell, B.R. Global survey of cell death mechanisms reveals metabolic regulation of ferroptosis. *Nature Chemical Biology* **2016**, *12*, 497–503.

(27) Gupta, P. B.; Onder, T. T.; Jiang, G.; Tao, K.;Kuperwasser, C.; Weinberg, R.A.; & Lander, E.S. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **2009**, *4*, 645-659.

(28) Xia, F.; Zuo, X.; Yang, R.; Xiao, Y.; Kang, D.; Vallee-Belisle, A.; … Plaxco, K.W. Colorimetric detection of DNA, small molecules, proteins, and ions using unmodified gold nanoparticles and conjugated polyelectrolytes. *PNAS* **2010**, *24*, 10837–10841.

(29) Sato, K.; Hosokawa, K.; Maeda, M. Rapid aggregation of gold nanoparticles induced by non-cross-linking DNA hybridization. *Journal of the American Chemical Society* **2003**, *27*, 8102-8103.

(30) Li, H.; Rothberg, L. Colorimetric detection of DNA sequences based on electrostatic interactions with unmodified gold nanoparticles. *PNAS* **2004**, *39*, 14036–14039.

(31) Wong, Siew Fang; Low, Kah Hin; Khor, Sook Mei. Differential-based biosensor array for fluorescence-chemometric discrimination and the quantification of subtle chloropropanols by cross-reactive serum albumin scaffolding. *Talanta*, **2020**, *218*, 121169.

(32) Wang, C.; Gong, B.; Bushel, P. R.; Thierry-Mieg, J.; Thierry-Mieg, D.; Xu, J., … Tong, W. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nature Biotechnology* **2014**, *32*, 926–932.

(33) Yoshioka, N.; Zangerl, B.; Phu, J.; Choi, A. Y. J.; Khuu, S. K.; Masselos, K.; … Kalloniatis, M. Consistency of Structure-Function Correlation Between Spatially Scaled Visual Field Stimuli and In Vivo OCT Ganglion Cell Counts. *Investigative Opthalmology & Visual Science* **2018**, *59*, 1693-1703.

(34) Terstappen, G. C.; Schlüpen, C.; Raggiaschi, R.; & Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nature Reviews Drug Discovery* **2007,** *6*, 891–903.

(35) Han, M. A.; Lee, D. H.; Woo, S. M.; Seo, B. R.; Min, K.; Kim, S.; … Kwon, T. K. Galangin sensitizes TRAIL-induced apoptosis through down-regulation of anti-apoptotic proteins in renal carcinoma Caki cells. *Scientific Reports* 2016, 6, 18642.

(36) Phosrithong, N.; Ungwitayatorn, J. Molecular docking study on anticancer activity of plant-derived natural products. *Medicinal Chemistry Research* **2009**, *19*, 817–835.

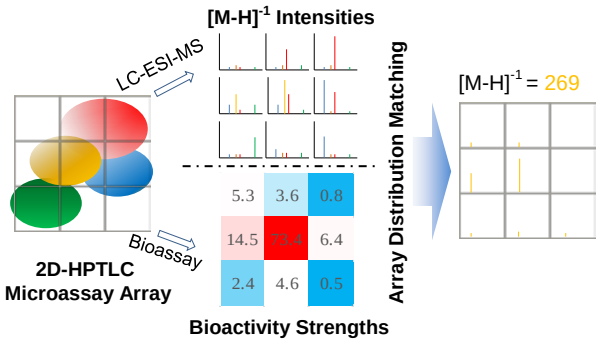
(37) Ferey, J.; Da Silva, D.; Lafite, P.; Daniellou, R. & Maunit, B. TLC-UV hyphenated with MALDI-TOFMS for the screening of invertase substrates in plant extracts. *Talanta* **2017**, *170*, 419–424.

(38) Rhee, I. K.; van de Meent, M.; Ingkaninan, K. & Verpoorte, R. Screening for acetylcholinesterase inhibitors from Am-aryllidaceae using silica gel thin-layer chromatography in combination with bioactivity staining. Journal of Chromatography A. 2001, 915, 217–223.

(39) Lee, M. S.; Kerns, E. H. LC/MS applications in drug development. *Mass Spectrometry Reviews* **1999**, *18*, 187 ± 279.

(40) Bu, X. Z.; Xiao, G. W.; Gu, L. Q.; Zhang, M. Chemical Study of Alpinia officinarum. Journal of Chinese *Medicinal Materials* **2000**, *2*, 84-87.

ToC graphic


[M-H]-1 Intensities
LC-ESI-MS
Bioassay
2D-HPTLC
Microassay Array
5.3
3.6
0.8
14.5
70.4
6.4
2.4
4.6
0.5
Bioactivity Strengths
Array Distribution Matching
[M-H]-1 = 269

Supporting Information

Coding Recognition of the Dose-Effect Interdependence of Small Biomolecules Encrypted on Paired Chromatographic-Based Microassay Arrays

Yifeng Deng*, Zhenpeng Lin, Yuan Cheng

*Corresponding author: dengyf@gdmu.edu.cn

S1 Preparation of the metal array grid interface

A design drawing of the metal array grid interface is shown (Figure S1). The metal array grid is made of 3D printed Co-Cr alloy (OEM according to drawings). The grid array format is the same as that of the 384-well microplate, with a format of 4.5 mm×4.5 mm square and horizontal 24×16 vertical arrays. The grid space is a regular inverted quadrangular pyramid. The vertical section of the grid frame is an isosceles triangle with a height of 2.5 mm. The width of the base of the isosceles triangle is the same as the wall thickness (approximately 0.70 mm) of the well of the 384-square-well microplate.

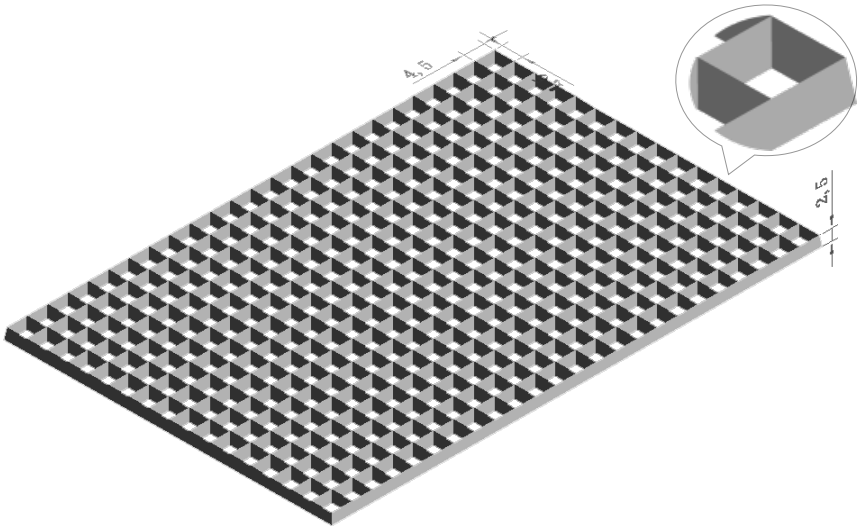


Figure S1. Metal array grid interface design

S2 Preparation of the square chip array of the silica gel layer

The silica gel layer of the chromatographic region of a sample's 2D-HPTLC fingerprint was split by the metal grid interface (S1) at a parallel pressure and stripped from a thin aluminum sheet into a flat square chip array. Specifically, the components of a sandwich composed of a silicone rubber sheet (thickness 2 mm), metal grid interface, PVDF membrane (Millipore, thickness 0.2 mm), dried HPTLC aluminum sheet and paper (PVDF film protection paper) were laminated in order from bottom to top. The sharp edges of the metal grid interface were aligned and faced the thin layer of silica gel across the PVDF membrane; then, the edges were pressed with a parallel pressure of about 8×10^3 kPa (Specac Atlas® Manual Hydraulic Press-15T). When the thin ductile aluminum sheet and the thin crisp layer of silica gel touched the sharp square opening edge of the grid interface with pressure, the thin layer of silica gel was split into square chips and was stripped from the thin aluminum sheet, thereby adhering to the PVDF membrane, which was slightly embedded in the metal grid interface. The stripping rate of the silica gel thin layer reaches 91.3%, and the chips of the silica gel thin layer are arrayed regularly and orderly as a square chip array.

S3 Preparation of the paired chromatographic-based microassay array

The openings of the 384-square-well filter plate (Pall 5072) were aligned with the array of silica gel layer chips adhered to the PVDF membrane (slightly embedded in the metal grid interface), and the PVDF membrane was touched with a 16-channel pipette (without the pipette tip) from the other side of the metal grid interface so that the silica gel layer chips fell into the corresponding wells of the 384-square-well filter plate (Figure S2). The components in the silica gel chips were eluted with methanol into another microplate. The microplate containing the eluent was evaporated to dryness on a vacuum centrifuge concentrator (269 ×g, 50°C, -80 kPa, 1.5 hours) and prepared as the stock chromatographic-based microassay array. The paired chromatographic-based microassay array with the same chromatographic distribution is prepared by transferring a portion of the sample from the stock

chromatographic-based microassay array to the corresponding array unit of another 384-well microplate.

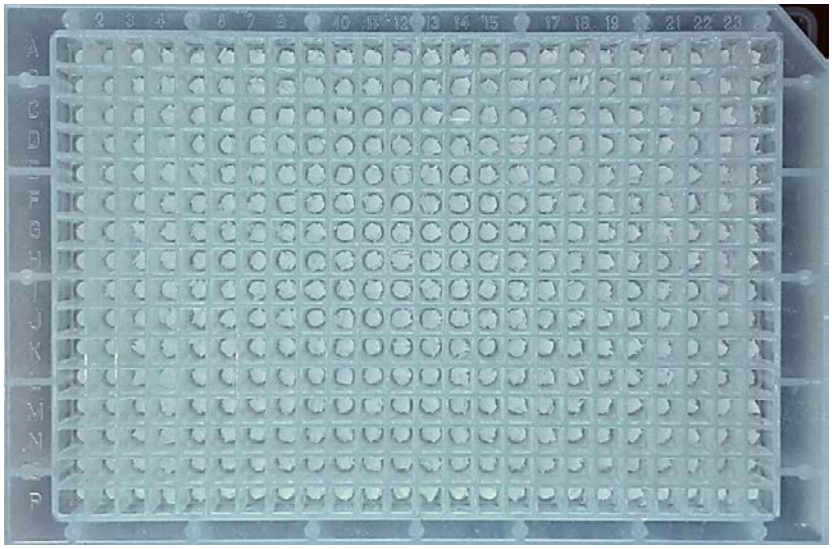


Figure S2. The components in the silica gel chips were eluted with a 384-square-well filter plate. The square chips of the silica gel layer were positioned into the corresponding 384-square-well filter plate

S4 Schematic diagram of preparation of chromatographic-based microassay array

A schematic diagram of preparation of Chromatographic-based microassay array is presented (Figure S3).

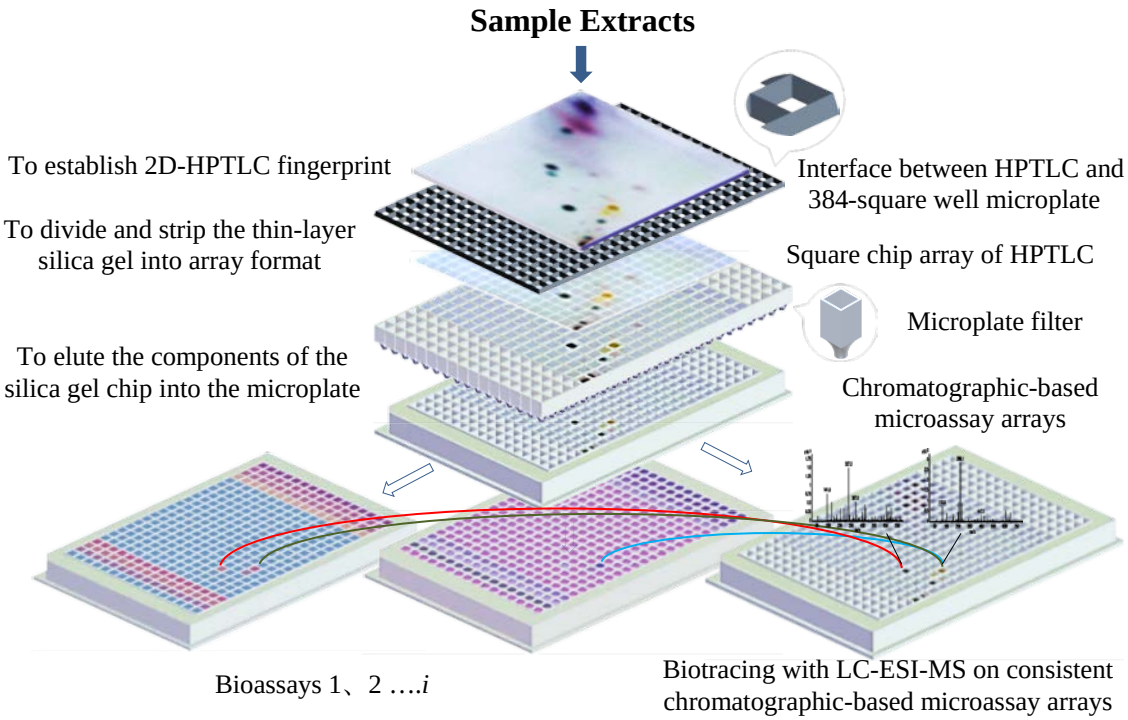


Figure S3. Schematic diagram of preparation of chromatographic-based microassay arrays