Video Article

Candidate Gene Testing in Clinical Cohort Studies with Multiplexed Genotyping and Mass Spectrometry

Sarah E. Ashley^{1,2}, Braydon A. Meyer^{2,3}, Justine A. Ellis^{2,3,4}, David J. Martino^{2,3,5}

¹Molecular Genetics of Chronic Inflammation and Allergic Disease, Max-Delbrück Center for Molecular Medicine

²Murdoch Childrens Research Institute

³Department of Paediatrics, University of Melbourne

⁴Centre for Social and Early Emotional Development, Faculty of Health, Deakin University

⁵Department of Paediatrics, University of Western Australia

Correspondence to: David J. Martino at david.martino@mcri.edu.au

URL: <https://www.jove.com/video/57601>

DOI: [doi:10.3791/57601](https://doi.org/10.3791/57601)

Keywords: Genetics, Issue 136, Genotyping, multiplex assay, mass spectrometry, MALDI-TOF, candidate gene, clinical cohort

Date Published: 6/21/2018

Citation: Ashley, S.E., Meyer, B.A., Ellis, J.A., Martino, D.J. Candidate Gene Testing in Clinical Cohort Studies with Multiplexed Genotyping and Mass Spectrometry. *J. Vis. Exp.* (136), e57601, doi:10.3791/57601 (2018).

Abstract

Complex diseases are often underpinned by multiple common genetic variants that contribute to disease susceptibility. Here, we describe a cost-effective tag single nucleotide polymorphism (SNP) approach using a multiplexed genotyping assay with mass spectrometry, to investigate gene pathway associations in clinical cohorts. We investigate the food allergy candidate locus *Interleukin13* (*IL13*) as an example. This method efficiently maximizes the coverage by taking advantage of shared linkage disequilibrium (LD) within a region. Selected LD SNPs are then designed into a multiplexed assay, enabling up to 40 different SNPs to be analyzed simultaneously, boosting cost-effectiveness. Polymerase chain reaction (PCR) is used to amplify the target loci, followed by single nucleotide extension, and the amplicons are then measured using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry. The raw output is analyzed with the genotype calling software, using stringent quality control definitions and cut-offs, and high probability genotypes are determined and output for data analysis.

Video Link

The video component of this article can be found at <https://www.jove.com/video/57601/>

Introduction

In human complex disease, genetic variants contribute to disease susceptibility and quantifying these variants may be useful for understanding pathogenesis, identifying high risk patient groups, and treatment responders. Indeed, the promise of precision medicine is dependent upon utilizing genomic information to identify patient subgroups¹. Unfortunately, within the complex disease biology space, where disease phenotypes are underpinned by substantial genetic heterogeneity, low penetrance, and variable expressivity, cohort size requirements for genome-wide approaches to identify novel candidates are often prohibitively large². Alternatively, a targeted candidate gene approach begins with an *a priori* hypothesis about specific genes/pathways in disease etiology³. Pathway analysis tools are commonly used to investigate the pathophysiology of an identified target loci, generating numerous candidate pathways to be explored. We demonstrate here a multiplexed genotyping approach allowing for the investigation of tens to hundreds of SNPs with one assay, suited to human cohort studies⁴. This approach is relatively high through-put, permitting hundreds to thousands of DNA samples to be genotyped for novel discovery studies and investigation of specific pathways. The methods outlined here are useful for identifying risk alleles and their associations with clinical traits in a relatively rapid and inexpensive manner. This platform has been highly advantageous for screening and diagnostic purposes^{5,6}, and more recently, for microbial infection⁷ and human papillomavirus⁸.

This protocol begins with selection of a set of genes for investigation, *i.e.*, the target regions, typically determined through literature searching, or *a priori* hypotheses for involvement in the disease process; or perhaps selected for replication as the leading associations of a discovery genome-wide association (GWA) study. From the gene set, the researcher will select a refined list of tag SNPs. That is, the linkage disequilibrium (LD), or correlation, amongst variants in the region is used to identify a representative 'tag SNP' for a group of SNPs in high LD, known as a haplotype. The high LD of the region means that the SNPs are often inherited together such that genotyping one SNP is sufficient to represent the variation at all SNPs in the haplotype. Alternatively, if following up on a definitive list of SNPs from many regions, *e.g.*, replication for a GWA study, this process may be unnecessary. For multiplexed genotyping, an assay must then be designed around these targets such that the amplification primers are of differing mass to those of the extension primers and products to produce interpretable mass spectra. These parameters are easily implemented by a multiplexed genotyping assay design tool. The forward and reverse primers from this design will be used to target the markers of interest and amplify the sequence containing the SNP. The extension primers attach directly proximal to the SNPs and a single, mass-modified, 'terminator' base that is complementary to the SNP is added. The terminator base prevents further extension of the DNA. The mass-modification of the base enables fragments differing by a single base to be detected by mass spectrometry. The plate containing the genotyping chemistry is then applied to a chip for measurement on a mass spectrometry platform. After applying appropriate quality controls

to the raw genotyping calls detected by the system, the data can be exported and used for statistical analysis to test for association with disease phenotypes.

Protocol

The genetic material used herein was ethically approved for use by the Office for Children HREC (Human Research Ethics Committee) (CDF/07/492), the Department of Human Services HREC (10/07) and the Royal Children's Hospital (RCH) HREC (27047).

1. Designing the Multiplexed Assay

1. Prepare a SNP list for assay design.
 1. Input the target region into the tagger function of Haploview (<https://www.broadinstitute.org/haploview/downloads>). Use the LD (correlation, r^2) between SNPs spanning the target region to generate a list of SNPs, which provides full coverage of the desired region with the fewest number of required variants. Generate a list of SNPs such that all alleles to be captured are correlated above a user-defined threshold (e.g., $r^2 \geq 0.8$).
2. Identification of suitable sequences for forward, reverse, and extension primers
 1. Enter the generated list of SNPs into the assay design tool of choice.
Note: The following instructions apply to the assay design tool used to generate the representative results, as specified in the **Table of Materials**.
 2. Once the assay design tool has been launched, select **New Genotyping Design**. Enter a name for the assay design into the **Design Name** field.
 3. Provide the SNP list generated in step 1.1.1 by clicking on the button labeled **Edit Text Input** with the instructions **rs or FASTA** to the left of the button.
Note: SNPs can be entered as a text list of reference (rs) SNP IDs or uploaded in FASTA format, or the SNP Group file format used by the tool. To upload files in these formats, select the **File Upload** button. The upper limit of variants that can be designed into a single assay, termed 'well,' is 40 SNPs. It is possible to specify SNPs that are a priority for the design before starting the run; these will be designed first. Control SNPs can be specified: for example, an assay to detect gender in case of sample mix ups or a control to determine that sufficient template was added if working with low or poor quality DNA. However, it should be noted that using control and priority SNPs can reduce the number of SNPs that can be designed into a single well.
 4. From the **Preset** menu, select the **High Multiplexing iPLEX Preset** option. From the **Organism** menu, select the organism from which the DNA to be genotyped derives. For the representative results presented here, this was Human.
Note: Mouse and bovine organisms are also compatible with the tool. One can also select **Other** if working with an alternative organism such as plant or microbe; however, the automated steps for finding proximal SNPs and optimal primer areas will not be available due to the absence of a reference genome.
 5. At the **Database** menu, select the latest build of the genome, or an alternative build if required, from the **Chemistry** menu. Select **iPLEX** and for the **Multiplex Level** field, enter the highest level of multiplexing: 40. Now, select **Begin Run**. "Step 1" of the tool will commence.
Note: During step 1, the tool retrieves and formats the SNP sequences. Here, errors might include formatting of the FASTA file in which case the FASTA file should be checked and reformatted as necessary. Another error is SNP rs numbers that have been merged with another rs number, in which case the rs number should be searched in a SNP database such as the NCBI's dbSNP to identify the new SNP rs number. Sequences around the SNP are searched for any known proximal SNPs that could interfere with the primer design. Potential PCR primer sites are compared against the genome to avoid non-specific amplification due to multiple hits in the genome.
 6. Await completion of "Step 2" of the design tool in which proximal SNPs will be identified.
Note: If using an organism that is not supported by this step, annotate the sequences with proximal SNPs in IUPAC annotation if this information is available. There are rarely errors during this step; however, errors may occur if the wrong reference organism or genome was selected, or if cDNA sequences were entered rather than genomic DNA. To optimize step 2, under **Advanced Settings**, it is possible to filter out SNPs based on a population if working within a specific population. Rare SNPs can also be filtered out by excluding those with a frequency below 1%. Finally, SNPs that do not have a validated status or do not have population information can also be excluded if desired.
 7. Await completion of "Step 3" of the design tool that identifies optimal primer regions. If preferred primer locations have been identified, or if working with a different organism, manually input primer locations by annotating the input sequence (the SNP group files) with the preferred primer regions in uppercase and the rest of the sequence in lower case, select **Preserve Case** from the Advanced Settings menu for this step.
Note: Errors for Steps 2 & 3 may include: (1) A proximal SNP blocking the design of a primer with the solution being to mask with a non-template base and order an oligo with a universal base like Isonine, or, design two primers, one with each of the alleles of the proximal SNP, or, design one primer with the common allele only. (2) Another common error is the failure to identify a unique site for the primer with the solution being to modify the amplicon length in order to identify a unique site. In the tool used for this paper, the default is 80 - 120 bp. This can be altered from 60 - 400 bp in the advanced settings dialogue under the Amplicon Length settings in "3. Identify Optimal Primer Areas" and also under the **Amplicon** tab of "4. Design Assays." Ensure both are changed. However, it should be noted that increasing the length of the amplicon when using degraded DNA is not recommended.
 8. Await completion of the assay design step ("step 4" of the design tool), here the tool aims to provide optimal pass separation between the extension primers and the alleles of all the SNPs in the multiplexed designs. After this step is complete, export the Oligo Order file, which is formatted for easy ordering through an oligo ordering vendor of choice.
Note: Amplification primers (Forward (F) and Reverse (R)) are designed to produce an optimal amplicon size of 80 to 120 bp. The amplification primer mass must differ from that of the extension primer and its product to avoid conflicting results in the mass spectrum and for overall PCR performance. The extension primer should be 15 to 30 nucleotides long (4,500 - 9,000 Da), designed so that the 3'

end will attach directly adjacent to the site of the polymorphism. All of these settings are automatically set in the tool used to generate the representative results.

- Export the design from the chosen tool and order primers from an oligomer provider: 25 nmol of each of the PCR primers (forward and reverse), and 100 nmol of each of the extension primers. See **Table 1** for the primers used to generate the representative results.

2. Genotyping (1 - 2 days)

- Prepare amplification primers.
 - If lyophilized primers were ordered, reconstitute with molecular grade water. Use a concentration of 100 $\mu\text{M}/\mu\text{L}$ for forward and reverse primers and 500 $\mu\text{M}/\mu\text{L}$ for extension.
 - Pool forward & reverse primers for each multiplexed assay into a stock primer mix, containing each primer at a concentration of 0.5 mM.
Note: For example, a 400 μL primer pool for 400 reactions:

$$\frac{0.5 \mu\text{M}}{100 \mu\text{M} \times 400 \mu\text{L}} = 2 \mu\text{L}$$
 2 μL of each primer, at a concentration of 100 $\mu\text{M}/\mu\text{L}$, is required for the stock primer pool. For a 30-plex assay with 60 forward and reverse primers, a total of 120 μL of primer would be required (the concentrated primer pool). 280 μL molecular-grade water would then be added to make a final volume of 400 μL of primer mix.
- Amplification using Polymerase Chain Reaction (PCR)**
 - Make a master mix for the PCR containing 100 nM of the primer mix detailed above, 500 μM of dNTPs, 2 mM of MgCl_2 , and 0.5 U of a DNA polymerase enzyme. 1 U is required for greater than 26-plex assays. Refer to **Table 2**.
 - Add 4 μL of this mix to each well of the 384-well PCR plate, along with 1 μL of genomic DNA at a concentration of 5 - 10 ng/ μL .
Note: Quality controls should also be included on the plate, such as a non-template control, and a positive control that has previously performed well with the assay (during a test run). If running multiple plates, multiple DNA samples from the other plate(s), e.g., in triplicate, should be run as technical controls for inter-plate variability.
 - Centrifuge the plate at 200 x g for 1 min at room temperature with a plate cover in place to ensure all liquid is contained at the bottom of the plate wells.
 - Run the PCR with the following thermocycler conditions: 4 min incubation at 94 $^{\circ}\text{C}$ to activate the DNA polymerase; 45 cycles of the following (denaturation at 94 $^{\circ}\text{C}$ for 20 s, annealing at 56 $^{\circ}\text{C}$ for 30 s, and extension at 72 $^{\circ}\text{C}$ for 1 min) and then a final extension at 72 $^{\circ}\text{C}$ for 3 min. Refer to PCR program in **Table 3**.
 - Perform agarose gel electrophoresis to ensure the DNA amplification is successful. Use 1 μL of PCR product (with 3 μL of loading buffer) on a 2% (w/v) agarose gel, made by mixing agarose powder and 0.5% Tris Borate EDTA (TBE) buffer, and run for 45 min at 100 V.
- Shrimp Alkaline Phosphatase (SAP) Purification Step**
Note: This step is used to remove unincorporated nucleotides. The SAP enzyme cleaves phosphate groups to prevent excess primer and unincorporated dNTPs from interfering in the upcoming extension reaction.
 - Make a master mix containing 1.7 units/ μL of Shrimp Alkaline Phosphatase (SAP) enzyme, 10x SAP buffer and make up to volume with molecular-grade water. Refer to **Table 4**.
 - Add 2 μL of the freshly made up SAP master mix to each well of the plate, already containing 5 μL from the PCR reaction. Centrifuge briefly and return to thermocycler at 37 $^{\circ}\text{C}$ for 40 min and then at 85 $^{\circ}\text{C}$ for 5 min for enzyme inactivation. Refer to **Table 5**.
- Primer Extension Reaction**
Note: The concentration of the extension primers in the extension primer mix must be adjusted by size (Linear Primer Adjustment – LPA). This is because there is an inverse relationship in peak intensity (on the mass spectra) and product mass. Adjustment of the extension primer concentration corrects for this relationship to produce peaks of equal intensities. Adjusted primers used to generate the representative results are provided in **Tables 6 & 7**. The volumes provided are to generate 1.5 mL of extension primer mix.
 - Calculate dilution factors for each primer using a gradient algorithm, available from Agena Biosciences ('Linear Primer Adjustment'), which adjusts for the mass and concentration of the primer, and the total number of primers in the multiplexed reaction (see **Tables 6 & 7**). Enter the UEP mass for each primer into the UEP_MASS field of the Linear Primer Adjustment sheet.
Note: The mass adjusted concentration for the primer will be automatically calculated in cell with the header text "Target Reaction Concentration (μM).". The volume of primer to be added to the primer pool will be output in the cell under the header text "Volume Stock Primer to be Added to Pool (μL).". The volume of water to be added to the primer pool to make up the calculated concentrations is output to the cell next to the text "Volume of PCR Grade RNase free H_2O to be added Primer Pool (μL).".
 - Take the volume of each primer as specified in the Linear Primer Adjustment file and make up a primer pool. Add the volume of water calculated by the algorithm to dilute the primer pool as determined in step 2.4.1.
Note: If the gradient algorithm method is unavailable, another method is to divide the primers into low mass and high mass and add double the concentration of the high mass primers relative to the low mass group. Three or four groups may be required for a high-plex assay (that is, more than 19 SNPs). However, this method will not yield as optimal call rates as the gradient algorithm method.
 - Assemble the extension master mix. For the low plex assay (1 - 18 plex), add 0.2 μL of buffer, 0.1 μL of termination mix, 0.94 μL of the LPA adjusted primer mix, and 0.0205 μL of enzyme (prepared on ice). For the high plex assay (19 - 36 plex), add double concentration of termination mix and enzyme (termination mix 0.2 μL and IPLEX enzyme 0.041 μL). Refer to **Table 8**.
 - Add 2 μL of the extension reaction master mix to the plate from the previous reaction, now bringing the total volume to 9 μL per well. Centrifuge briefly and place in the thermocycler: 94 $^{\circ}\text{C}$ for a 30 s initial incubation, 40 cycles of 1 x 94 $^{\circ}\text{C}$ for 5 s with 5 x (52 $^{\circ}\text{C}$ for 5 s, followed by 80 $^{\circ}\text{C}$ for 5 s), and then 72 $^{\circ}\text{C}$ for 3 min. Refer to **Table 9**.
- Resin purification.**

Note: Excess salts can now be removed from the reaction with the use of resin.

1. Add 16 μ L of deionized water to the wells of the retrieved plate, bringing the volume in the plate up to 25 μ L.
2. Apply resin, usually 6 mg of dimple resin plate (purchased separately), evenly to the whole plate. Leave the resin to dry in the resin dimple plate for 20 min at room temperature before applying to the reaction plate. After addition of the resin, rotate plate for 5 min at room temperature either by hand or with a suspension mixer at a slow speed (e.g., 7.5 rpm).

Note: Prior to loading the genotyping analyte onto the genotyping chip, the samples and the assay layout must be input into the software interface. This can be achieved using the Design Summary file from the assay design tool.

6. Measurement on a mass spectrometry platform.

1. Centrifuge the plate at 3,200 x g for 5 min to avoid application of resin to the mass spectra chip.
2. Upload the plate layout of samples to the system's software interface prior to the run.
Note: Step 2.6.3 should only be performed by trained personnel.
3. Apply the genotyping analyte mixture from the plate to the genotyping chip with a dispensing system. Next, load the chip onto the MALDI-TOF system to generate the mass spectra from the analyte mixture, which can then be interpreted with the aid of the platform's software interface.

Note: The system generates the mass spectra by directing short pulses of UV light at each 'pad,' corresponding to a single well of the genotyping plate, of the mass spectra chip. This pulse results in desorption and ionization of the analyte. These ionized DNA molecules are then propelled to the top of the vacuum tube by a high voltage electrostatic field, separating out lighter ions, which travel faster and reach the detector first, from the heavier ions. The "time of flight" of these analytes are recorded by the system, from which the mass of each DNA fragment can be determined and thus the nucleotide base present at the SNP can be inferred.

3. Genotyping Data

Note: Quality control and data interpretation are not the focus of this methodology paper. However, the following will briefly cover interpretation of the mass spectra.

1. Manual inspection and quality control of mass spectra.

Note: The raw mass spectra are generated by the system and analyzed with the software as listed in the **Table of Materials**. A Gaussian mixture clustering method is applied to this output to make probabilistic genotyping calls.

1. Open the analysis software. Select **Typing Analyzer**. From the **File** menu, select **Open Wells from File** and select the xml file with the spectrum data generated in step 2.6.3. The chip name for the run should be visible, select it and a "traffic light" pane of the 384-well plate will be displayed with a list of the SNPs in the selected well. From here, select **Call Cluster Plots** to view a scatterplot of data points for each SNP.
Note: It is recommended that a 'no call' result be applied for low intensity SNPs. In the software used for the representative results, a clustering magnitude cut-off of 5, which is higher than default, is suggested for stringent quality control. This setting can be adjusted in the **Magnitude Cutoff** field under **Parameters** in the **Post Processing Clusters** pane. Select **Autocluster** from the **Tools** tab to run a cluster analysis.
2. Perform manual inspection of genotype calls by examining the Cartesian Plots in the Post Processing Pane. In the Assay pane, select the SNP of interest and click on the **Post-Processing Clusters** tab, a cluster plot with genotype clustering will be visible with sample IDs and corresponding genotypes for the SNP.
3. Examine the genotype call clusters and assess how well each individual call clusters with the other calls for the SNP. The software provides some boundary lines on the plot for each genotype for guidance; see example of a genotype calls cluster plot from an *IL13* SNP (**Figure 1**). If the call does not cluster tightly with other calls and sits clearly outside of a cluster, or if it has very low intensity, right click the datapoint and select **Change Call** to manually set it to **No Call**.
4. To ensure high quality genotyping data is used for downstream analysis, exclude SNPs and individuals with call rates below a QC threshold, e.g. 90%. Once appropriate adjustments have been made to the call data, export the data for statistical applications. In the software used to generate the representative results (listed in the **Table of Materials**) this is achieved with the **Plate Data** selection from the view menu and selecting **Save As**.

Representative Results

With the protocol described above, we genotyped tag SNPs across the Th2 immune gene *IL13* in a cohort of food allergy cases and controls⁹. We applied logistic regression analysis, adjusted for ancestry and other potential covariates, to test whether the genetic variants within the region of interest increased food allergy risk. **Table 10**⁹ shows that one variant rs1295686 is associated with challenge-proven food allergy and we confirmed this association in a replication cohort using the same multiplexed genotyping approach. A meta-analysis of the results from the two cohorts provided strong evidence of association of *IL13* with FA (**Table 11**)⁹. We genetically inferred and adjusted for ancestry in the analysis using an ancestry informative SNP marker panel, adapted from a published panel¹⁰. We genotyped the panel using the same multiplexed assay approach.

The associated variant, rs1295686, has been previously identified as an asthma risk loci¹¹ and associated with other allergic immunological parameters¹², suggesting it may be a general allergic disease risk loci. The next steps would be to conduct further studies, such as differential expression analysis, mapping physical interactions with a chromatin capture method, fine mapping of the region, and haplotype analysis, to pinpoint the functional variant and characterize the biology behind the observed association with food allergy.

Figure 1: Cartesian cluster plot for the *IL13* SNP rs1295686. Yellow clusters represent homozygous genotype calls for the A allele, blue for the G allele, and green for heterozygous calls. The few genotype calls that did not cluster with either the homozygous or heterozygous mass spectra were set to "no call."

Table 1: Primer sequences used to generate the representative results for *IL13*. The Name column contains the SNP ID, the well number (as mentioned previously, the "well" number refers to the ID of the multiplexed assay), and the type of primer (F for forward, R for reverse, and UEP for the extension primer). The next columns contain the sequence of the primer, the size of the primer and the purification type. It should be noted that *SPINK5* and an Ancestry Informative Markers (AIM) panel were genotyped using the same assay although these results are not discussed with the presented representative results. SNPs ending in _CEU and _CHB refer to AIM SNPs for the Northern Europeans from Utah and the Han Chinese in Beijing, China populations from the 1000 genomes project respectively. [Please click here to download this file.](#)

Master Mix		Low Plex (≤26 SNPs)		High Plex (>26 SNP)	
		× 1	× 200	× 1	× 200
Reagent	Conc. in 5 µL				
Water (deionized)		1.9	380	1.8	360
Buffer	1 × (2 mM MgCl ₂)	0.5	100	0.5	100
MgCl ₂	2 mM	0.4	80	0.4	80
dNTPs	500 µM	0.1	20	0.1	20
Primer mix**	100 nM	1	200**	1	200**
Taq polymerase	0.5 U / 1 U	0.1	20	0.2	40
Total		4 µL	800 µL	4 µL	800 µL

**already diluted with deionized H₂O as detailed in 2.1.2

Table 2: Reagents and concentrations required to make the Amplification PCR primer mix. The master mix volumes are given for 200 reactions because 400 (enough to cover a 384-well plate) will not fit into a 1.5 µL tube and thus 2 × 200 should be made in separate tubes. Volumes specified under Low Plex should be used if the multiplexed assay "well" contains less than or equal to 26 SNPs, if greater than 26 SNPs are in the well use the High Plex volumes.

Thermocycler conditions
94 °C for 4 min
45 cycles of (94 °C 20 s, 56 °C 30 s, 72 °C 1 min)
72 °C for 3 min
4 °C hold

Table 3: Thermocycler conditions for the Amplification PCR.

Master Mix	× 1	× 410
Water (deionized)	1.53	627.3
10× buffer	0.17	69.7
SAP enzyme (1.7 U/μL)	0.3	123
Total	2 μL	820 μL

Table 4: Reagents and concentrations for the master mix for the SAP reaction. Master mix volumes are given for 410 reactions to cover a 384-well plate with a margin for pipetting error.

Thermocycler conditions
37 °C for 40 min
85 °C for 5 min
4 °C hold

Table 5: Thermocycler conditions required for the SAP reaction.

#	Extension Primers	Original stock concentration (μM)	Misc.	UEP_ MASS	Target Reaction Concentration (μM)	EXT Primer Pool Concentration (μM)	Volume Stock Primer to be Added to Pool (μL)
1	rs36110_CHB	500		4359.8	0.560	5.36	16.09
2	rs10488619_CHB	500		4546	0.602	5.76	17.29
3	rs1986420_CEU	500		4761.1	0.648	6.21	18.62
4	rs2934193_CEU	500		4964.3	0.690	6.61	19.82
5	rs4968382_CEU	500		5047.3	0.707	6.77	20.30
6	rs1227647_CEU	500		5136.4	0.724	6.93	20.80
7	rs1402851_CEU	500		5338.5	0.763	7.30	21.91
8	rs4705054	500		5476.6	0.788	7.55	22.64
9	rs6928827	500		5678.7	0.824	7.89	23.68
10	rs9325072	500		5762.8	0.839	8.03	24.10
11	rs4841401_CHB	500		5853.9	0.855	8.18	24.55
12	rs11098964_CHB	500		6052.9	0.888	8.50	25.51
13	rs2416504_CHB	500		6174	0.908	8.69	26.08
14	rs1860933	500		6264.1	0.923	8.83	26.50
15	rs2193595_CHB	500		6391.2	0.943	9.03	27.08
16	rs1519260_CHB	500		6469.2	0.955	9.14	27.43
17	rs11184898_CHB	500		6588.3	0.973	9.32	27.95
18	rs679832_CEU	500		6757.4	0.998	9.56	28.68
19	rs326626_CEU	500		6765.4	1.000	9.57	28.71
20	rs1612904	500		6873.5	1.015	9.72	29.17
21	rs862942	500		6960.5	1.028	9.84	29.53
22	rs2486448_CHB	500		7169.7	1.058	10.13	30.38
23	rs4824001_CEU	500		7341.8	1.081	10.35	31.06
24	rs6552216_CEU	500		7465.9	1.098	10.51	31.54
25	rs1488299_CHB	500		7620	1.119	10.71	32.13
26	rs1347201_CHB	500		7626	1.119	10.72	32.15
27	rs315280_CHB	500		7747	1.135	10.87	32.60
28	rs11203006_CHB	500		7834.1	1.146	10.97	32.92
29	rs4240793_CHB	500		8036.3	1.172	11.22	33.66
30	rs9325071	500		8219.4	1.194	11.43	34.30
31	rs12678324_CEU	500		8394.5	1.215	11.64	34.91
32	rs4653130_CEU	500		8455.5	1.223	11.71	35.12
33	rs9275596	500		8537.6	1.232	11.80	35.39
34	rs12595448_CHB	500		8603.6	1.240	11.87	35.62
Volume of PCR Grade Rnase free H ₂ O to be added Primer Pool (μL)							561.78

Table 6: Well 1 adjusted primers used to generate the representative results. A table of extension primers adjusted by size, including the target concentration, the volume used in the primer pool, and final concentration for each primer in the primer pool for well 1. The volume of water required to make up to the final volume is provided at the bottom of the table. The total volume of the primer pool was 1.5 mL.

#	Extension Primers	Original stock concentration (μM)	Misc.	UEP_ MASS	Target Reaction Concentration (μM)	EXT Primer Pool Concentration (μM)	Volume Stock Primer to be Added to Pool (μL)
1	rs10515597	500		4482.9	0.588	5.63	16.89
2	rs2759281_CEU	500		4921.2	0.681	6.52	19.57
3	rs17641748	500		5080.3	0.713	6.83	20.48
4	rs1698042_CEU	500		5201.4	0.737	7.05	21.16
5	rs5753625_CHB	500		5411.6	0.776	7.43	22.30
6	rs3912537_CEU	500		5811.8	0.848	8.12	24.35
7	rs6141319_CEU	500		5883.8	0.860	8.23	24.70
8	rs1002587_CEU	500		6026.9	0.884	8.46	25.39
9	rs1295686	500		6172	0.908	8.69	26.07
10	rs16877243_CEU	500		6386.2	0.942	9.02	27.05
11	rs1103811	500		6595.3	0.974	9.33	27.98
12	rs6510332_CEU	500		6790.4	1.003	9.61	28.82
13	rs6595142_CHB	500		6874.5	1.016	9.72	29.17
14	rs4484738_CEU	500		6966.5	1.029	9.85	29.55
15	rs1432975	500		7028.6	1.038	9.94	29.81
16	rs10879311_CEU	500		7317.8	1.078	10.32	30.97
17	rs864481	500		7416.9	1.092	10.45	31.35
18	rs1295687	500		7447.8	1.096	10.49	31.47
19	rs12644851_CHB	500		7634	1.120	10.73	32.18
20	rs4265409_CHB	500		7926.2	1.158	11.09	33.26
21	rs2927385_CHB	500		7997.2	1.167	11.17	33.52
22	rs1538956_CHB	500		8286.4	1.202	11.51	34.54
Volume of PCR Grade Rnase free H ₂ O to be added Primer Pool (μL)							899.42

Table 7: Well 2 adjusted primers used to generate representative results. A table of extension primers adjusted by size, including the target concentration, the volume used in the primer pool, and final concentration for each primer in the primer pool for well 2. The volume of water required to make up to the final volume is provided at the bottom of the table. The total volume of the primer pool was 1.5 mL.

Master Mix	Low Plex (1-18)		High Plex (19-36)	
Reagent	× 1	× 410	× 1	× 410
Water (deionized)	0.7395	303.195	0.619	253.79
Buffer	0.2	82	0.2	82
Termination mix	0.1	41	0.2	82
Primer mix (LPA Adjusted)	0.94	385.4	0.94	385.4
iPLEX enzyme*	0.0205	8.405	0.041	16.81
Total	2 μL	820 μL	2 μL	820 μL

Table 8: Reagents and concentrations for the extension reaction master mix. In this case the volumes for Low-Plex reactions should be used if there are 18 or fewer SNPs in the well. For 19 or more SNPs use the High Plex reactions volumes. This is different to the SNP requirements of low-plex and high-plex of the amplification PCR master mix detailed in **Table 2**.

Thermocycler conditions
94 °C for 30 s
40 cycles of (94 °C 5 s, (5 cycles of 52 °C 5 s, 80 °C 5 s))
72 °C for 3 min
4 °C hold

Table 9: Thermocycler conditions for the extension reaction

Food allergic cases vs non-food allergic controls					
SNP	A1	P	OR	L95	U95
rs1295686	A	0.003	1.75	1.2	2.53
rs2243297	A	0.13	2.1	0.8	5.51
rs1295687	G	0.19	1.63	0.78	3.41
rs2243211	A	0.22	1.45	0.8	2.64
rs1295683	T	0.25	1.32	0.82	2.13
rs2243248	G	0.51	1.21	0.69	2.11
rs2243300	T	0.51	1.19	0.7	2.02
rs1800925	T	0.6	1.11	0.75	1.63
rs3091307	G	0.62	1.1	0.75	1.6

Table 10: Variant rs1295686 of the *IL13* locus is associated with challenge-proven food allergy. Logistic regression analysis, adjusted for ancestry, sex and presence of atopic eczema, revealed that variant rs1295686 is associated with challenge proven food allergy in the discovery cohort (n = 367 cases and 156 non-allergic controls). The SNP column lists the markers being examined, the A1 column lists the minor allele, used as the reference allele for the association test. The P column lists the P-values from the regression analysis, the OR column lists the corresponding odds ratios and L95 and U95 are the lower and upper limits of the 95% confidence interval from the analysis. Data reproduced from Ashely *et al.*, 2017⁹.

A. Replication analysis: food allergic cases vs non-food allergic controls						B. Meta-analysis – Discovery and Replication					
SNP	A1	OR	L95	U95	P	P	P(R)	OR	OR(R)	Q	I
rs1295686	A	1.37	1.03	1.82	0.03	0.0005	0.0006	1.5	1.5	0.31	3.32
rs1295687	C	1.1	0.68	1.78	0.7	0.3	0.3	1.24	1.24	0.38	0

Table 11: Replication in an independent population confirms association between *IL13* variant rs1295686 and challenge-proven food allergy. Logistic regression, corrected for ancestry principal components, sex and atopic eczema, in an independent population (n=203 food allergic cases and 330 non-allergic controls) confirmed the association between variant rs1295686 and food allergy. Meta-analysis was then conducted, providing the highest level of evidence for an association between the SNP and outcome, with little evidence of heterogeneity in effect sizes between the two populations at this SNP. Here the columns are as per **Table 10**, with additional P(R) and OR(R) values for the random effects meta-analysis model vs the fixed effects model (P and OR). Q and I are the P-value for the Cochran test and the I-index respectively, both are measures of the heterogeneity in effect sizes between the studies. Data reproduced from Ashely *et al.*, 2017⁹.

Discussion

Here, we demonstrate the method of multiplexed genotyping using mass spectrometry. The representative results were generated using PCR paired with MALDI-TOF mass spectrometry⁴ with assay chemistry listed in the **Table of Materials**¹³. With this platform, we generated a total of 11,295 genotypes on 1,255 individuals for 9 SNPs within 40 h in the lab.

We illustrate the usefulness of the technique in answering genetic hypotheses by generating and analyzing SNP data for the candidate gene *IL13* in a discovery clinical cohort phenotyped for food allergy and with independent replication. The platform is relatively robust to lower DNA quality and requires minimal starting material, an input of only 10 ng. Critical steps in the protocol include the following: careful troubleshooting of the assay design process to ensure maximal SNP inclusion in the multiplexed assays, successful amplification of the target regions during the amplification PCR, successful single nucleotide extension during the extension reaction, and loading of the assay and sample plate layout into the analysis software in order for the software to accurately assign the mass spectra data.

As previously mentioned, the assay design tool is compatible with the use of mouse and bovine polymorphisms, in addition to human. For other organisms such as microbe or plant, "other" can be selected in the "Organism" menu. However, the automated steps for finding proximal SNPs and identifying optimal primer areas will not be available.

During the design process, if a proximal SNP is blocking the identification of an optimal primer area, one can either remove the SNP from the design and select another in high LD (e.g., $r^2 = 1$), or choose to design two primers, one with the reference allele of the SNP and one with the alternative allele, or designing one primer with the common allele, or masking the SNP with a universal base (e.g., Inosine). If there is an issue

with identifying a unique site for the primer, one can change the amplicon length settings to find a unique site. The default setting is 80 - 120 bp but this can be modified from 60 - 400 bp. However, it should be noted that if working with degraded DNA (e.g., from FFPE tissue), it is not ideal to increase the amplicon length.

Errors during the assay design step of the tool might include high hairpin or primer dimer potential. Both settings can be relaxed to try an accommodate the designs. However, it is recommended that thresholds are not relaxed below 0.8; start with 0.9 to see if this setting is sufficient to rescue the SNPs and reduce to 0.8 only if necessary. In the advanced settings, it is possible to change the number of design iterations (up to 10) under the **Multiplex** tab of "4. Design Assays", as well as the best iteration selection criteria to either **Highest Average Multiplex** or **Fewest Rejects by Low Plex**. Increasing the number of design iterations can be advantageous because the tool will take the first SNP in the order they were provided and will design an assay. Then it tests if the next SNP is compatible for the same multiplex assay. Thus, the order of the SNPs matter. With the multiple iterations option selected, the software will randomize the order of SNPs and try other iterations. This may increase the number of SNPs able to be designed into each multiplex or may decrease the number of SNP rejects. However, it will also come at a time cost, as the design tool will take much longer at this step.

Multiplexed genotyping is a quick and affordable approach for investigation of the loci of interest. The method is streamlined and allows the running of approximately 760 samples of up to 40-plex SNPs in 10 hours, permitting tens of thousands of genotypes to be generated in less than one day¹⁴. The mass spectra can be easily analyzed with tools provided by the platform.

Some restrictions to the platform include an estimated loss of 5 - 10% of SNPs that will fail the assay design process. This can sometimes be corrected by selection of an alternative tag or proxy SNP. The Broad Institute's SNP Annotation and Proxy Search (SNAP) database is one such tool that facilitates identification of proxy SNPs¹⁵. Another limitation of the system is that it is a targeted platform and therefore does not allow for extensive novel discovery, as achieved with genome-wide approaches. Furthermore, as the approach uses tag-SNP proxies to maximize coverage across genes, fine-mapping studies may subsequently be needed to determine the precise causal variant.

Multiplexed genotyping is still a useful platform for the interrogation of disease associated SNPs identified in GWA study approaches or hypothesis driven candidate and pathway gene exploration. Future applications for the platform are likely to be novel uses for diagnostics and screening in fields such as cancer and clinical virology. Overall multiplexed genotyping with mass spectrometry is a quick, cost-effective and reliable medium-throughput method for genotyping candidate loci.

Disclosures

The authors have nothing to disclose.

Acknowledgements

The authors have no acknowledgements.

References

1. Aronson, S.J., Rehm, H.L. Building the foundation for genomics in precision medicine. *Nature*. **526** (7573), 336-342, (2015).
2. Ball, R.D. Designing a GWAS: power, sample size, and data structure. *Genome-Wide Association Studies and Genomic Prediction*. 37-98, (2013).
3. Kwon, J.M., Goate, A.M. The candidate gene approach. *Alcohol research and health*. **24** (3), 164-168, (2000).
4. Oeth, P., Mistro, G.D., Marnellos, G., Shi, T. and van den Boom, D. Qualitative and quantitative genotyping using single base primer extension coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MassARRAY). *Single Nucleotide Polymorphisms: Methods and Protocols*. 307-343, (2009).
5. Pusch, W., & Kostrzewa, M. Application of MALDI-TOF mass spectrometry in screening and diagnostic research. *Current pharmaceutical design*. **11** (20), 2577-2591, (2005).
6. Su, K. Y., et al. Pretreatment epidermal growth factor receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *Journal of clinical oncology*. **30** (4), 433-440, (2012).
7. Singhal, N., Kumar, M., Kanaujia, P. K., & Viridi, J. S. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in microbiology*. **6**, 791, (2015).
8. Cricca, M., et al. High-throughput genotyping of high-risk Human Papillomavirus by MALDI-TOF Mass Spectrometry-based method. *New Microbiologica*. **38** (2), 211-223, (2015).
9. Ashley, S., et al. Genetic Variation at the Th2 Immune Gene *IL13* is Associated with IgE-mediated Paediatric Food Allergy. *Clinical & Experimental Allergy*. **47** (8), 1032-1037, (2017).
10. Bousman, C.A., et al. Effects of NRG1 and DAOA genetic variation on transition to psychosis in individuals at ultra-high risk for psychosis. *Translational psychiatry*. **3** (4), e251, (2013).
11. Moffatt, M. F., et al. A large-scale, consortium-based genomewide association study of asthma. *New England Journal of Medicine*. **363** (13), 1211-1221, (2010).
12. Granada, M., et al. A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. *Journal of Allergy and Clinical Immunology*. **129** (3), 840-845, (2012).
13. Oeth, P., et al. iPLEX assay: Increased plexing efficiency and flexibility for MassARRAY system through single base primer extension with mass-modified terminators. *Sequenom application note*. 27, (2005).
14. Ellis, J.A. and Ong, B. The MassARRAY System for Targeted SNP Genotyping. *Genotyping: Methods and Protocols*. 77-94, (2017).
15. Johnson, A. D., et al. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. **24** (24), 2938-2939, (2008).