

MassARRAY® System for Somatic Mutation Screening Designing custom assays for highly sensitive and accurate analysis

Somatic mutations, long implicated as a driver of abnormal cell proliferation, do not occur randomly in cancer. Mutations are found more frequently in certain genomic regions and/or cellular pathways. Genetic profiling assesses DNA samples for the purposes of identifying people with a higher likelihood of experiencing a harmful side effect to a drug, or developing an inherited disease as the result of a specific genotype. More specifically, somatic mutation profiling delves into non germline aberrations associated with oncogenesis and proliferation. Mutation profiling has the potential to identify quickly which signaling pathways have been co-opted to drive the proliferation of a particular tumor type. Somatic mutation profiling of individual patients may also be a useful tool in guiding therapeutic options. The challenge in analyzing somatic mutations involves detecting tumor cell specific mutations in a background of wild-type genotypes from adjacent non-tumor cells.

MALDI-TOF MS has been proven as a highly advantageous method for somatic mutation profiling^{1,2}. A group of clinical researchers at the Dana Farber Cancer Institute designed a somatic mutation panel to enable large scale screening of cancer samples with increased clinical sensitivity. This panel, comprised of 238 oncogene mutations in 17 oncogenes, was used to screen >1,000 human tumor samples from 17 types of solid tumors. Relevant mutations were confirmed in 30% of the samples. More interestingly, novel mutations were detected that had not been previously reported due to the higher sensitivity of this method. For example, *EGFR* (Epidermal Growth Factor Receptor) mutations were detected with as little as 9% of the total DNA population³. Another study highlights the use of the platform for screening 400+ mutations in >30 oncogenes and tumor suppressors against ~900 clinical samples of which >35% had at least one of the mutations⁴.

Sequenom developed a panel for somatic mutation profiling, OncoCarta™ Panel v1.0, which enables routine and robust processing of tumor samples. This panel allows any researcher to analyze 238 mutations in 19 oncogenes, including some of the more challenging genes of interest such as *EGFR*, *KRAS*, and *BRAF* with as little as 10% mutation frequency⁵.

While the OncoCarta™ Panel features a comprehensive set of mutations suitable for most research projects, individual groups may have specific interests beyond the OncoCarta™ panel. This application note outlines special considerations required for somatic mutation assay design and analysis utilizing Sequenom chemistries with MALDI-TOF MS detection so users can expand on Sequenom's panel or develop their own panels.

Basic Principles for Somatic Mutation Assay Design

In standard SNP genotyping, samples typically follow a Mendelian inheritance pattern and will be classified as either homozygous (100%) or heterozygous (50%) for a given allele (**Figure 1A**). Assay designs are straightforward, there is no preference whether a particular allele is of low or high mass since the mass signals are clearly distinct (e.g., 0%, 50%, or 100%). Assaying samples for somatic mutations requires detection of mutations in functional amounts (**Figure 1B**). The most common form of DNA sample assayed for somatic mutations is DNA derived from a mixture of tumor tissue and healthy tissue. The tumor tissue may harbor a somatic mutation, and depending on the proportion of tumor tissue to healthy tissue in the sample there could be a far lower mutation frequency compared to germline mutations. The task of designing assays is further complicated by the complexity of the genetic aberrations and the properties of the surrounding sequences. The MassARRAY® system provides a superior method for screening low percentages of mutated DNA against a background of wild-type DNA.

Figure 1—the difference in detecting and analysis of genotyping versus somatic mutations

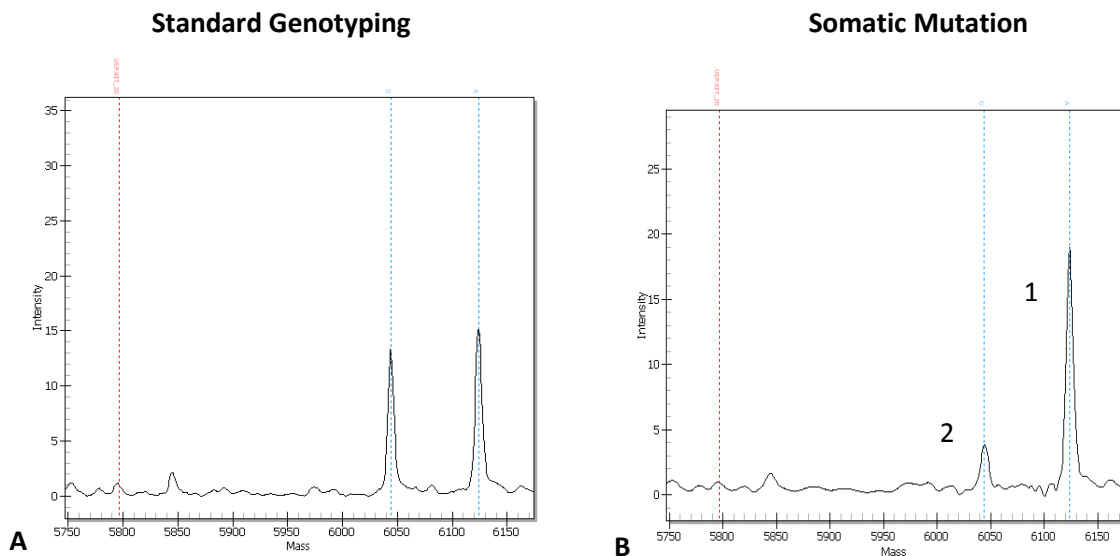


Figure 1 (A) Spectrum shows two peaks corresponding to the two different alleles with an approximate frequency of 50%, indicative of a heterozygous genotype. **(B)** Spectrum shows a peak corresponding to a wild-type allele (1) and a mutant allele (2) at much lower frequency.

This quantitative readout is beyond what Sanger sequencing can provide. The MassARRAY® system excels at detecting low abundance mutations. One study describes an assay design that detects <3% mutant allele for BCR-ABL⁶. Assays can be designed to any number of target regions and screened in parallel. This offers added flexibility beyond sequencing, which interrogates a specific gene or the whole genome.

Improving Sensitivity for Low Frequency Mutations

Assay performance and sensitivity are largely dependent on the signal-to-noise ratio of the mass signals within an assay. A systematic approach to somatic mutation assay design and analysis is necessary to reduce the likelihood of overlapping signals, and to detect low abundance signals. Following a set of rules for design and subsequent analysis will improve the discrimination of potential false positives from true positives. Key design criteria include:

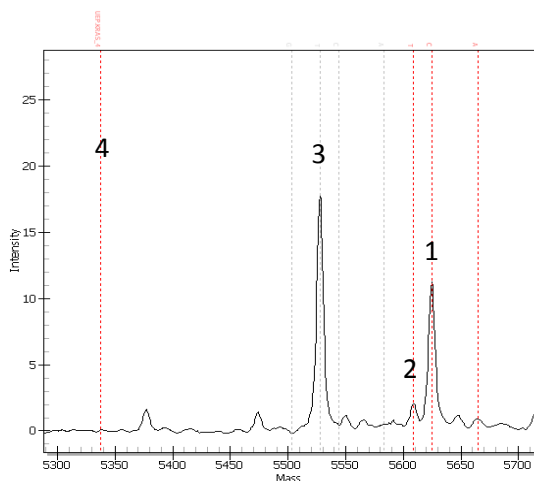
- Mass spacing and positioning of alleles (e.g., wild-type vs. mutant) for higher sensitivity
- Reducing the effects of salt and matrix adduct contribution in MALDI-TOF MS
- Thoughtful multiplexing of assays

The MassARRAY® system detects any nucleic acid product in the 4,500-9,500 daltons mass range. Each MS peak is called a mass signal. Every mass signal has a lower limit of detection. A limiting factor is the influence of other mass signals within the spectrum that may interfere with accurate interpretation. Close proximity of neighboring mass signals (non baseline resolved mass signal pairs) can overlap with the desired signal. It is essential to reduce interference in order to measure low frequency mutations and to improve assay sensitivity. The following section describes different types of interference.

Analytes within the Assay

Mass signals in an assay include unextended primer, analytes of interest (e.g. extension products for wild-type and mutant alleles), and salt or matrix adducts (**Figure 2**).

Figure 2—typical profile of a single assay for somatic mutation analysis

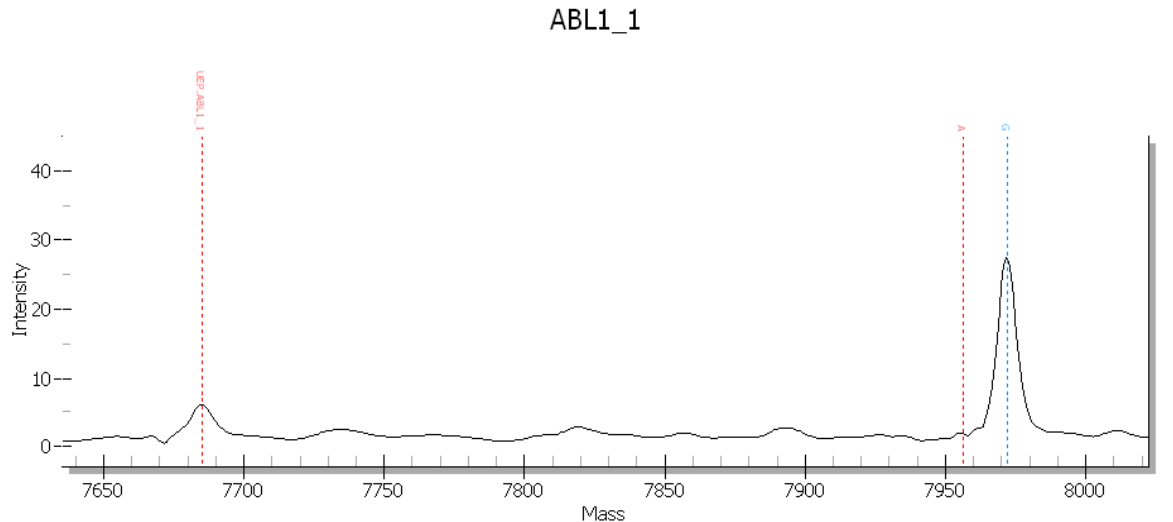


Spectrum at left shows a peak corresponding to a wild-type allele (1) and a mutant allele (2) at a much lower frequency. This reaction is multiplexed and contains a peak (3) from an additional assay. Low abundance peak (2) could be a potential adduct of (3) or a real mutation. The unextended primer (4) is depicted by a red dashed line at approximately 5,340 Da (4). The vertical lines depict the expected mass signals in the assay. For more information on adducts, see adduct section.

Assays need to be optimally designed to avoid interference from other mass signals. Each mass signal should be distinct to ensure accurate identification and separation. It is important to avoid interference from analytes and/or adducts within an assay.

Figure 3 represents one of the simplest designs with a single mutation being investigated. The mutant peak (A) has a lower mass than the wild-type peak (G), preventing adduct interference. Optimal assay design is discussed in the section *Designing assays for a single mutation*.

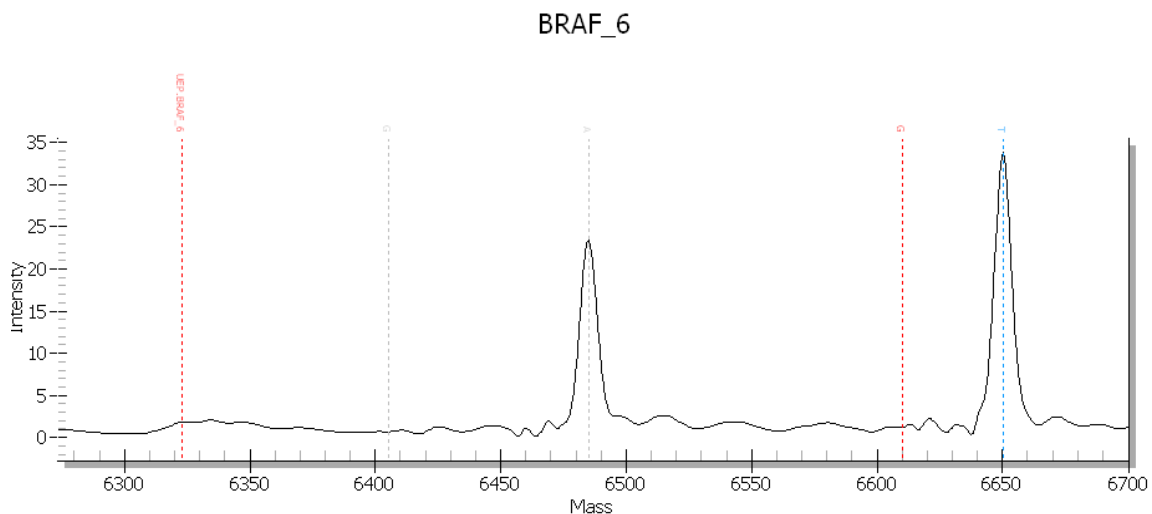
Figure 3—a single mutation assay for *ABL1* optimally designed



Analytes in a Multiplex assay

Multiplexing assays within the same well increases the number of mass signals in the spectrum (**Figure 4**).

Figure 4—typical profile of a multiplexed assay



The vertical lines show the expected mass signals in the *BRAF* assay. The mutant G allele (6610 Da) has no call, whereas the wild-type T allele (6650 Da) shows 100% abundance. It is ideal for the unextended primer (UEP) to have minimal signal intensity (6325 Da).

The spectrum (**Figure 4**) also shows an assay multiplexed with another assay. The mutant allele was designed at 6405 Da, and the wild-type allele was designed at 6485 Da. The multiplex still produces a spectrum with distinct mass signals that do not interfere with the other calls. The two assays are well separated from one another. In general, reducing the multiplexing to ~8-10 assays per well, from the standard 40 SNPs for genotyping, improves the sensitivity and performance for each assay, as the assays are spaced to avoid interference.

Designing multiplexed assays with mutations from the same gene can be more complex. If the gene has multiple mutations within close proximity, one could design a single PCR primer set for the amplification. If the same PCR primer set is used to evaluate a target region, the most straightforward multiplexed assay would have extension primers of varying length to ensure mass difference. This approach saves on primer costs and reduces the number of wells that need to be run per sample. However, the assays need to be designed with care unless different primer extension lengths are used. Extension primers with nearly identical sequences of the same length will be of similar mass. While there may be cost savings in the run, the sensitivity could be compromised.

For highly polymorphic regions, design assays in different multiplexes. The *EGFR* gene, for instance, has a large number of hot spots with >50 somatic mutations. The OncoCarta™ Panel contains 24 multiplexes to encompass this region. The sensitivity may have been reduced if only a few multiplexes were designed as the analyte mass signals are of similar mass. Multiplexes may also be designed with different target regions. Complex mutation analysis works well with the MassARRAY® system as very little DNA (10-20 nanograms) is required per reaction, and the running costs per reaction for the system are highly affordable.

Adducts—an introduction to contaminant peaks

Small molecules attach to extension products and alter their mass to form adducts. Adducts are always present in spectrum to some degree. They are observed as a group of small peaks at regular intervals of increasing mass relative to any large extended or unextended primer peaks. The most commonly observed adducts are salt adducts (**Table 1**).

Table 1—Adducts observed in MassARRAY® spectrum

| Adduct | Mass Shift (Da) | Frequency | Potential Issue |
|-----------------|-----------------|----------------------|-------------------|
| Ammonium | 18 | Rare | A/G |
| Sodium | 22 | Very common (10-20%) | C/A |
| Potassium | 38 | Common (5-10%) | C/G or G/T |
| Manganese/ Iron | 55 | Rare | A/T |
| Copper | 63 | Rare | Multiplexed assay |
| Matrix 1 | 135 | Rare | Multiplexed assay |
| Matrix 2 | 189 | Common | Multiplexed assay |

Adducts increase the level of noise, potentially interfering with small mass signals that may represent low frequency mutations. The most common adducts are sodium (+22 Da) and potassium (+38 Da). The mass shift for sodium can overlap with a C to A mutation (+24 Da). For potassium, the mass shift can overlap with a G to T mutation (**Figure 5A**) and a C to G mutation (+40 Da). Matrix adducts and multiple salt adducts can be more easily mistaken for a genuine mutant peak due to the larger mass shift. These peaks add a larger additional mass to the extend product that can overlap with an expected mutant peak for a neighboring assay with higher mass (**Figure 5B**).

Figure 5—Examples of adducts interfering with mutation interpretation

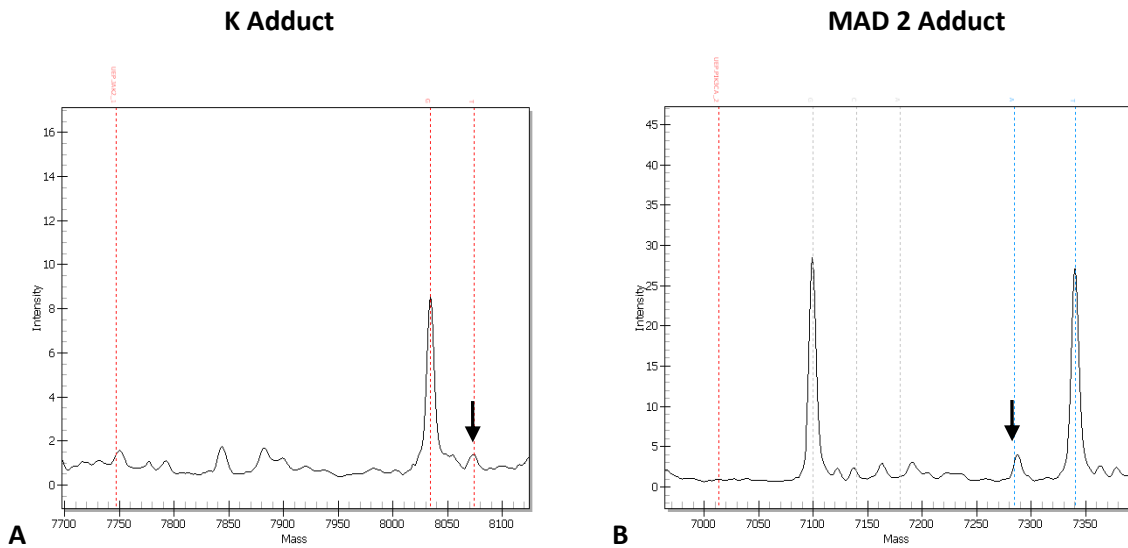


Figure 5 (A) Potassium Adduct (+38 Da of the wild-type peak) overlapping with a mutant T peak (+40 Da of the WT peak). (B) A MAD2 adduct (+189 Da) from a neighboring assay overlaps with an expected mutant peak. The vertical lines depict expected mass signals in the assay. The arrows point to interference that obscures the desired signals.

Figure 6 further illustrates potassium adduct interference (+38 Da) where adducts cannot be differentiated from the mutant alleles (+40 Da). The TYPER™ software reports a ~6% mutation frequency, however the sample was a known control with 100% wild-type DNA. This specific assay has been designed to detect a C/A mutation using the reverse strand (G/T with a +40 Da separation). However, designing against the forward strand (C/A with a +24 Da separation) would result in interference by a sodium adduct (+22 Da). Neither direction is ideal for this particular sequence as the mutant will always be the high mass allele. Guidelines for the preferred design strand are provided in the next section *Overcoming adduct interference*. Use of SpectroCHIP™ II is recommended to reduce potassium interference.

Figure 6—example of potassium adduct interference

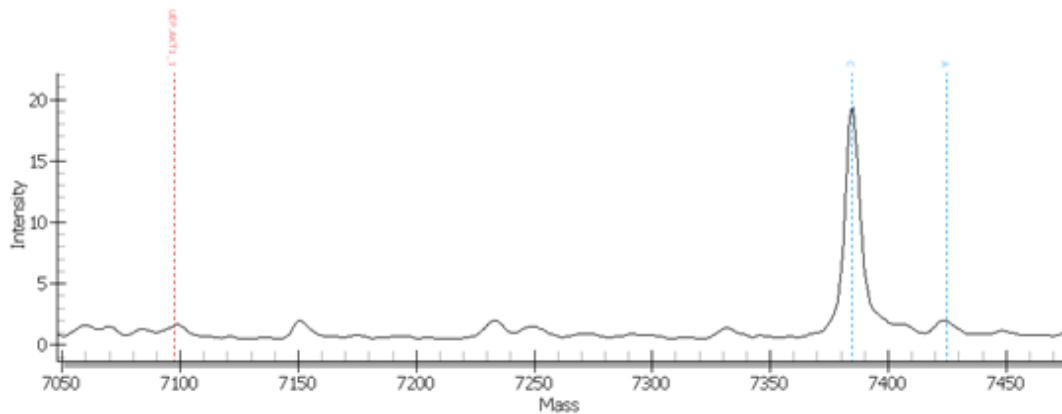


Figure 6 An example of a less optimal assay design where the mutant A allele (7425 Da) is obscured by a potassium adduct (+38 Da).

Overcoming adduct interference

The foremost rule is to select a design where the mutant allele will have a smaller mass than the wild-type allele. This reduces the likelihood of adduct interference (**Figure 5**). **Table 2** provides guidelines highlighting the preferred strand for assay design (shaded in green). The preferred strand is selected based on the preference for either a lower mass mutant allele, a larger mass separation, or selection of a design that results in a potential potassium adduct vs. a sodium adduct. The sodium adducts occur more frequently than potassium adducts (**Table 1**).

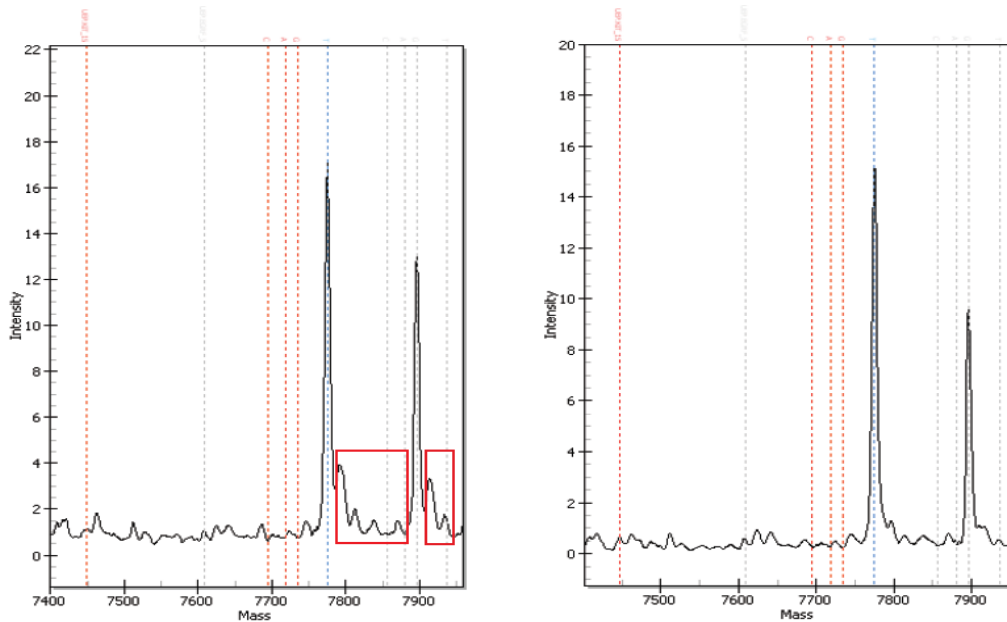
Table 2—Preferred strand based on specific base changes

| Forward | | | Reverse | | |
|---------|-----|-----------------|---------|-----|-----------------|
| WT | MUT | Mass Difference | WT | MUT | Mass Difference |
| A | C | 24 | T | G | 40 |
| A | G | -16 | T | C | 80 |
| A | T | -56 | T | A | 56 |
| C | A | -24 | G | T | -40 |
| C | G | -40 | G | C | 40 |
| C | T | -80 | G | A | 16 |
| G | A | 16 | C | T | -80 |
| G | C | 40 | C | G | -40 |
| G | T | -40 | C | A | -24 |
| T | A | 56 | A | T | -56 |
| T | C | 80 | A | G | -16 |
| T | G | 40 | A | C | 24 |

SpectroCHIP® II reduces adduct contribution

Use of SpectroCHIP® II will help to reduce some of the common adduct contribution. In a recent test, SpectroCHIP® II was shown to decrease salt adduct formation up to ~20% compared to the original chip. Potassium adducts are particularly reduced (**Figure 7**).

Figure 7—SpectroCHIP® II reduces salt adducts and improves signal to noise for low abundance mutation analysis



A *KIT* mutation run on the original SpectroCHIP® (left) vs. SpectroCHIP® II (right). Salt adducts (red boxes) are significantly reduced when the assay is run on SpectroCHIP® II.

While the mass window is large (4,500-9,500 Da), the analytes in a single assay are typically very close together. Analytes are extension products ranging from 17-30 nucleotides. However, for a single mutation, the wild-type and mutant alleles are the same length and only differ by a single base change. The resultant extension products may differ by only 16 Da as is the case for an A/G mutation. To accurately distinguish mutations, it is important to adhere to the following key assay design guidelines.

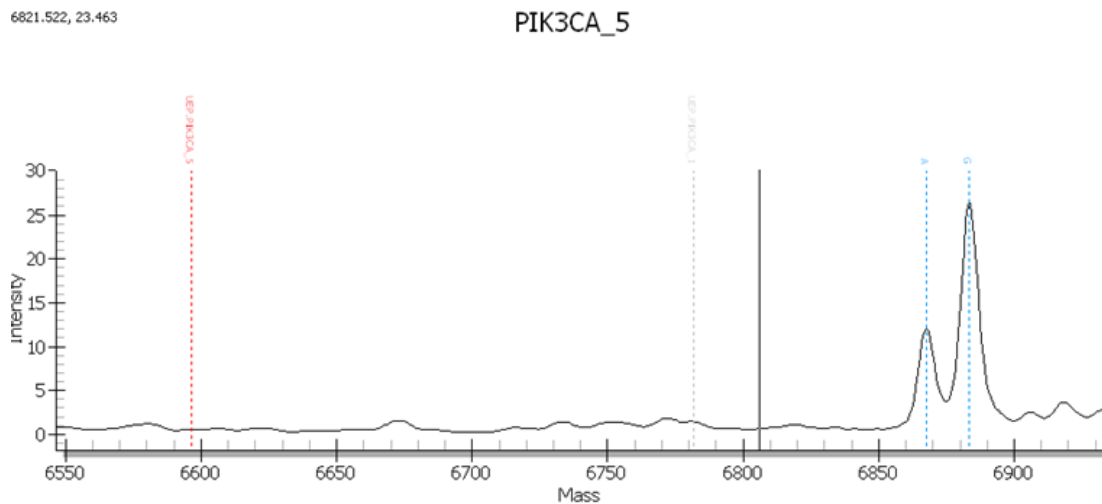
Ideal practice for assay designs:

- The mutant allele should have a lower mass than the wild-type allele
- Consider salt and matrix adducts and design to avoid interference from Na⁺/K⁺ ions

A low abundance mutation can be difficult to discern if there is interference from the wild-type allele. To measure the allele frequencies accurately, the assay should be designed so that the mutant allele has a lower mass than the wild-type allele. This is imperative for mutations with small mass differences, but is also best practice for any somatic mutation assay design. This

guideline is important because wild-type sequences typically have higher abundance in most samples. An assay for PIK3CA E542K was designed with a low mass mutant allele (**Figure 8**). The large percentage of wild-type allele (G) could have compromised detection of the mutant allele (A) if the assay was designed with the wild-type of lower mass given the close proximity of the mass signals (16 Da). The assay detected 70% wild-type and 30% mutant allele abundance.

Figure 8—an ideal assay design #1



Designing assays for multiple mutations

More complex assay designs involve the detection of more than one mutant allele or multiple base mutations. For instance, a multiple allele mutation C<A<G has one wild-type allele (C) and two mutant alleles (A and G). The ideal assay would have the wild-type allele with the highest mass, and the mutant alleles of lower mass. If such an assay is not feasible, each mutation can be investigated independently. In general, the following guidelines should be considered:

1. Design the assay so that the wild-type allele has the highest mass (**Figure 9**). Typically, the forward strand is used as a starting point.
2. If the mutant allele(s) have a higher mass than the wild-type allele, try to design the assay using the reverse orientation.
3. In the event that the wild-type allele cannot be designed with the highest mass in either orientation, try to ensure that at least one or as many of the mutant alleles as possible are of lower mass than the wild-type allele (**Figure 10**). The more mutants that can be designed with lower mass, the better the assay.
4. If the alleles overlap within an assay, try to design extension primers of different lengths or multiplex the mutation into different wells. This is discussed in the *Combined Assay* section.

Figure 9—Example of an ideal assay

7794.074, 24.272

FLT3_2

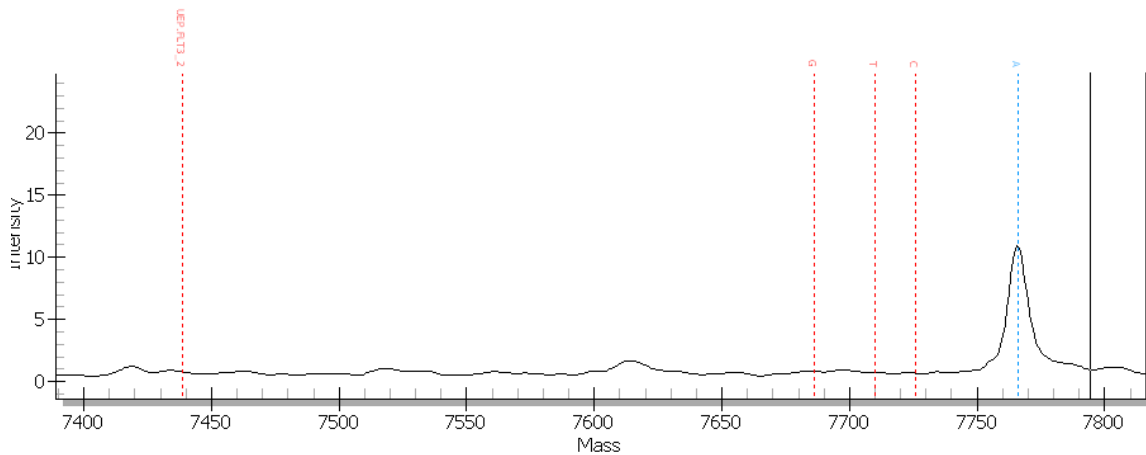
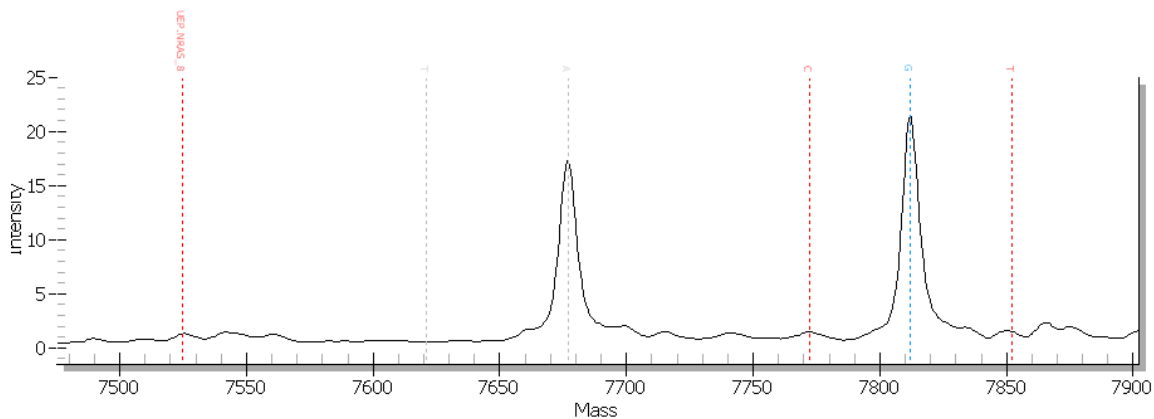


Figure 10 shows an assay that is less optimal as the wild-type allele falls between the two mutations. However, there is still adequate mass separation to make an accurate call. Some manual interpretation of the T allele may be required for abundance less than 10%, due to the close proximity between the G /T mutation (40 Da) and the potassium adduct (38 Da). In this situation, use of SpectroCHIP® II is highly recommended to reduce potassium adducts.

Figure 10—Example of an acceptable assay



Challenging Somatic Mutations

A somatic mutation may sometimes be undetectable using standard assay design. This can occur with a substitution, insert or deletion where there is a common nucleotide.

Wild-type: GTC ATC TTA GGG CTT CAA **T**AT CTT AAG CGG TTC C
 Mutant 1: GTC ATC TTA GGG CTT CAC **C**AT CTT AAG CGG TTC C
 Mutant 2: GTC ATC TTA GGG CTT CAA **G**AT CTT AAG CGG TTC C

Consider designing extension primers from the forward strand using standard assay design:

Wild-type: GTC ATC TTA GGG CTT CAA **T**AT CTT AAG CGG TTC C
 Mutant 1: GTC ATC TTA GGG CTT CAC **C**AT CTT AAG CGG TTC C
 Mutant 2: GTC ATC TTA GGG CTT CAA **G**AT CTT AAG CGG TTC C

The resultant extension products are identical for the wild-type and mutant allele 2, making it impossible to differentiate the mutation.

Wild-type: GTC ATC TTA GGG CTT CAA (Forward design)
 Mutant 1: GTC ATC TTA GGG CTT CAC (Forward design)
 Mutant 2: GTC ATC TTA GGG CTT CAA (Forward design)

Tip 1—Design the Assay using the Reverse Strand

Assays can sometimes be designed successfully using the reverse, complementary strand. All three alleles can now be readily distinguished.

Wild-type: G GAA CCG CTT AAG AT**A** (Reverse design)
 Mutant 1: G GAA CCG CTT AAG AT**G** (Reverse design)
 Mutant 2: G GAA CCG CTT AAG AT**C** (Reverse design)

This example illustrates an assay design that has been overcome by changing the orientation.

Tip 2—Design Multiple Assays

In the next example, a mutation can be detected, but it is not clear which mutation is present within the sample.

Wild-type: GTC ATC TTA GGG CTT CAA **C**AT CTT AAG CGG TTC C
 Mutant 1: GTC ATC TTA GGG CTT CAC **C**AT CTT AAG CGG TTC C
 Mutant 2: GTC ATC TTA GGG CTT CAC **G**AT CTT AAG CGG TTC C

Multiple assays can be combined in order to fully resolve individual mutations. The forward design will distinguish Mutant 1 from the wild-type allele, but cannot differentiate between the two mutations. The resultant extension products are identical for mutants 1 and 2.

Wild-type: GTC ATC TTA GGG CTT CA**A** (Forward design)
 Mutant 1: GTC ATC TTA GGG CTT CA**C** (Forward design)
 Mutant 2: GTC ATC TTA GGG CTT CA**C** (Forward design)

Using a combination of forward and reverse strands fully resolve the issue, so that specific mutations can be identified, not only that a mutation is present.

Wild-type: GTC ATC TTA GGG CTT CA**A** (Forward design)
 Mutant 1: GTC ATC TTA GGG CTT CA**C** (Forward design)
 Mutant 2: G GAA CCG CTT AAG AT**C** (Reverse design)

Tip 3—Use hME chemistry instead of single-base extension

Homogeneous mass extension (hME) is a multiple base primer extension. Its primary benefit arises when designing assays with sequences that have short repetitive regions. The enzyme, Thermostable Sequenase, and nucleotide mixes can extend either to the SNP site, or past the SNP site, with one extension primer design.

For the sequence below:

Wild-type: GTC ATC TTA GGG CTT CA**A** **T**GG GTT TTT CGG TTC
 Mutant 1: GTC ATC TTA GGG CTT CA**C** **C**GG GTT TTT CGG TTC
 Mutant 2: GTC ATC TTA GGG CTT CA**A** **G**GG GTT TTT CGG TTC

The hME extension primer (underlined) is designed as follows:

Wild-type: GTC ATC TTA GGG CTT CA**A** **T**GG GTT TTT CGG TTC
 Mutant 1: GTC ATC TTA GGG CTT CA**C** **C**GG GTT TTT CGG TTC
 Mutant 2: GTC ATC TTA GGG CTT CA**A** **G**GG GTT TTT CGG TTC

The resultant extension products are as follows:

Wild-type: GTC ATC TTA GGG CTT CA **AT** (Forward design)
 Mutant 1: GTC ATC TTA GGG CTT CA **C** (Forward design)
 Mutant 2: GTC ATC TTA GGG CTT CA **AG** (Forward design)

In this example, a nucleotide termination mix of dA/ddC/ddG/ddT is used. For additional details, see the *Multiplexing Homogeneous MassEXTEND™* application note.

Combined Assays

Combined assays involve more than one assay in order to make a mutation call with certainty. These types of assays are particularly necessary for a region with many mutations, or a region with complex mutations such as multiple base changes or larger insertions and deletions.

As an example, the oncogene Kirsten ras sarcoma virus (KRAS), is frequently implicated in a number of cancers, and contains a number of mutations within codon 12. In order to cover this region, we designed a series of extension primers in two different assays. KRAS mutations are interpreted from results of the following assays in the OncoCarta™ Panel v1.0.

Table 3—Design example of Combined Assays for Complex Region

| KRAS | Assay 1 + 2 | Assay 1 | Assay 2 |
|------|-------------|---------|---------|
| WT | CC | C | C |
| G12V | AC | A | C |
| G12A | GC | G | C |
| G12D | TC | T | C |
| G12C | CA | C | A |
| G12S | CT | C | T |
| G12R | CG | C | G |
| G12F | AA | A | A |

Several basic guidelines should be considered for highly complex regions:

1. If one mutation is close to another mutation, different extension primers can be used with the same set of PCR primers as long as the extension primers do not overlap with any of the mutations or with any other primers (i.e., at least 20 Da).

- a. As an example, the sequence below:

GTC ATC TTA GGG CTT C**[A/C]**G TAT CTT AAG CGG TTC CAT **[G/C]**AG AAA CTC GGA CAG

The two underlined sequences could be used as extension primers. The software will automatically calculate the extension primer masses.

- b. It is important to remember that extension primers need to be at least 15 nucleotides long or heavier than 4,500 Da in order to be detected by the MassARRAY® System.
2. If the sequences are fairly close to one another, it may be helpful to design one extension primer in the forward orientation, and the other extension primer in the reverse orientation.

In the sequence below:

5' GTC ATC TTA GGG CTT C[C/G]G ATA GGG GAC ACC GCA TA[A/C]CTT AAG CGG TTC CAT GAG 3'
 3' CAG TAG AAT CCC GAA G[G/C]C TAT CCC CTG TGG CGT AT[T/G]GAA TTC GCC AAG GTA CTC 5'

Design the first mutation (C/G) using the forward strand as the wild-type G allele will have a higher mass than the mutant C allele (40 Da). Design the second mutation (A/C) from reverse orientation, as the wild-type A allele will be +80 Da from the mutant G allele. See **Table 2** as a guideline.

The resultant extension primers in this example will be:

Sequence 1: GTC ATC TTA GGG CTT C

Sequence 2: CTC ATG GAA CCG CTT AAG

3. If a change in orientation does not help to generate a large enough mass difference between the two extension products, the mutations should be designed in individual multiplexes. This increases the number of wells per sample, but also enables detection of complex mutations.

Paralogous genes

Paralogs are two genes or gene clusters at different chromosomal locations in the same organism that have structural similarities indicating that they may have derived from a common ancestral gene and have since diverged from the parent copy by mutation and selection or drift. Assays should be designed to avoid simultaneous detection of paralogs otherwise the mutation frequency may be an aggregate of the two genes or clusters of genes. It is important to identify sequence differences between the gene of interest and its paralog, and use these differences to design target-specific assays. If sequence differences cannot be found, the assay will potentially detect both paralogs resulting in reduced sensitivity.

Primer hairpin or stem-loop structures

Primer hairpins occur when two regions of the same molecule, usually a nucleotide sequence palindrome, base-pair to form a double helix that ends in an unpaired loop. Hairpins prevent access and primer extension for mutation detection. In the following sequence, the forward extension primer will generate a strong hairpin resulting in the incorporation of a C base at single base extension (in this example *C is the mutation).



The following recommendations may reduce hairpin formation:

1. Add a non-templated base at a nucleotide adjacent to the stem-loop structure. In this example, the sequence is altered to TTT CCG [T]AA AGG AGA CTT TC[T/C]. In this example, the non-templated base can be a T, A, or C
2. Note that hairpin formation may still occur, however the mismatch at the 3' base of the hairpin will suppress single base extension from the hairpin. The mismatched base will result in reduced annealing; however at the -14 base position from the 3' end, this is minimal.

Dimerization

Dimerization occurs when two identical molecular entities interact to form a single dimer. Certain extension primer sequences may form dimers. In the following sequence, ACC TGG CGT GGA CCA GG[G/T], the forward extension primer will generate a strong self dimer resulting in the incorporation of a T base single base extension (in this example C is the mutant).

```
ACC TGG CGT GGA CCA GG*
          ||||
GGA CCA GGT GCG GTC CA
```

The following recommendations may reduce dimerization:

1. Change the sequence at the 5' end to a non-templated base. In this example, the nucleotide in the second position is altered to A[T]C TGG CGT GGA CCA GG[G/T]. The non-templated base can be either T, A, or G.
2. Primer dimer formation may still occur, however the mismatch at the 3' base of the primer will suppress single base extension from the dimer. The mismatched base will result in reduced annealing; however at -16 bases from the 3' end, this will be minimal.
3. The above examples are intended as a guide only and may not result in a working assay. Empirical testing is essential, as these modifications may result in lower yield and annealing success.

Assay Validation

Assays should be validated and evaluated for their analytical sensitivity. Key criteria for evaluating assay quality include testing with control samples for primer extension efficiency, allele ratios of wild-type vs. mutant DNA, testing with synthetic oligonucleotides (if no appropriate control DNA is available), and/or using statistical methods to evaluate assay performance.

Testing with control samples

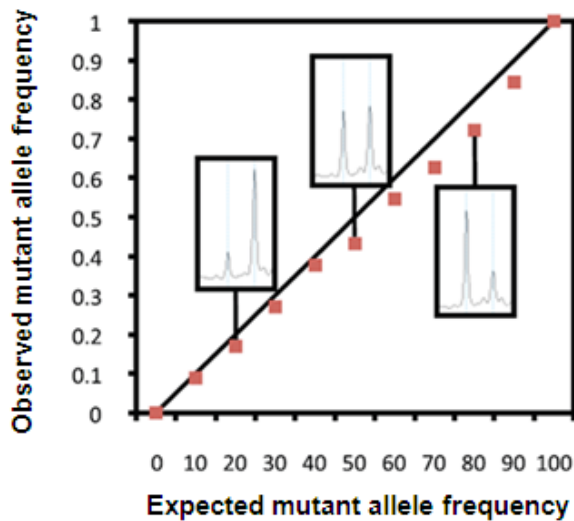
Assay validation for somatic mutation detection is not trivial. Frequently, control samples from actual tumor tissue or cell lines are not readily available. This is particularly likely if the mutations are rare. However, a significant number of quality metrics for an assay can be assessed with wild-type DNAs (i.e. genomic DNA from cell lines or healthy individuals) before using more valuable sample sources. Among the important metrics is the primer extension yield (amount of primer extended into the desired products), which is indicative of the PCR amplification efficiency and the post-PCR primer extension efficiency. A good assay should exhibit primer extension yields of $\geq 80\%$ for the amount of input genomic DNA used. The selection of an appropriate cut-off and testing of the robustness of this metric can further be supported by performing dilution experiments, in which the input DNA is decreased to the point where the reaction fails, e.g. the primer extension yield drops below the desired cut-off. Assays should be evaluated with control material where available. For example, this can be genomic DNA derived from cell lines harboring a mutation in heterozygous state.

In addition to the primer extension cut-off, the allele ratio between mutant allele and wild-type allele can be used as a quality metric and should center around 50% for a confirmed heterozygous sample derived from a cell line. If experiments are performed in replicates, the precision of this measurement can be evaluated. At sufficient DNA concentrations high enough to not introduce a sampling error, these semi-quantitative measurements should be precise within $\pm 3\%$. Deviations from the expected 50:50 ratio can be caused by imbalanced allele amplification, or, as discussed earlier, by co-amplification of paralogous regions carrying the wild-type allele. If such heterozygous genomic DNA is available, mixtures with wild-type cell lines can be prepared to further study the linearity of the assay and determine the limit of detection for the mutant allele in a mixture.

Testing with synthetic oligonucleotides

If no control material is available, synthetic templates can be designed and ordered with the mutation of interest from an oligonucleotide vendor. For the initial assessment, single-stranded oligonucleotides of the amplicon length harboring the same PCR primer and extend primer sequences are sufficient (~80-100 bases). For longer term use, plasmids cloned with the target sequences are preferable and serve as effective assay controls. This approach was used to evaluate the OncoCarta™ Panel (**Figure 11**).

Figure 11—Analytical sensitivity of an OncoCarta™ assay



Genomic DNA was mixed with a synthetic oligonucleotide containing a mutation ranging from 0 to 100%. This type of experiment can help determine whether the assay will perform optimally, or whether another round of assay design is required prior to running against samples. Multiple rounds of design and optimization may be required to ensure ideal performance of a somatic mutation assay.

Synthetic oligos may be a potential contamination risk resulting in false positive calls. To reduce the risk of contamination in the lab, order synthetic oligos in liquid format rather than dried-down or lyophilized. This will reduce the amount of contamination due to aerosol and pipetting issues during resuspension. Set up the oligos in a separate hood from the PCR primers. Sequenom also recommends ordering synthetic oligos either after the PCR primers have been ordered, or from a different vendor.

Statistical evaluation of assay performance

For suboptimal assays (e.g., assays where it is not possible to avoid adduct overlap) one should run a sufficient number of samples in order to develop criteria for a normal distribution. Statistical methods for outlier detection can then be applied. In brief, the tightness of the allele clustering is evaluated, and mutations are identified if they present as outliers from the normal cluster, which is assumed to be a wild-type call. If an assay shows very tight allele clusters for the wild-type allele, it can still be a useful assay. More details are described in the next section.

Mutation detection as outlier detection

Genotyping calls in TYPED are generally determined based on static parameters derived from spectral properties (e.g., static thresholds for call probability categories – conservative, moderate, aggressive). This approach is simple to interpret and works well in many cases. There are cases, however, for which such static thresholds are not appropriate – for example, assays which show strong allele bias. For these cases, a very successful method has proven to be one of calculating genotyping by means of modeling the allele ratios as a mixture of normal distributions (using the principle of Expectation Maximization, EM). This method, documented in

the TYPER user manual as post-processing clustering, uses the arc tangent of the allele ratio as a random variable for fitting a mixture of 3 normal distributions, using EM. Genotyping calls are then made by identifying the mode (component) of this mixture for which a given allele ratio has the highest probability. Even this method has its limitations: all three populations (low-mass homozygote, heterozygote, high-mass homozygote) should be well represented in each plate/chip for which this mixture is to be evaluated. If this requirement is not met, as in the case of assays meant to capture rare mutations, the problem of assigning a call can be approached as an outlier detection problem. This type of approach has been used for example for targeted CNV genotyping⁸. Several statistical techniques are available for these analyses⁹.

Data Analysis

Using mass spectrometry, each DNA species is represented by a signal in the mass spectrum. The ratios between the intensities of the mass signals are reflective of the ratio of the different DNA species. Specificity is limited by the ability to accurately determine if a signal that should represent mutated DNA is actually caused by mutated DNA and not by a reaction artifact. Because of the complexity of the multiplexed mass spectra, this task can be challenging at times. The *OncoCarta Manual Analysis Guide* provides guidelines to perform mutation identification by breaking the process down into several smaller tasks. The aim is to learn and to perform manual analysis of OncoCarta results confidently, leading to results with a higher degree of accuracy. This can also be applied to other somatic mutation data.

Summary

Using these guidelines for assay design provides a better understanding of how to generate primer sequences for somatic mutations. This will enable the design of assays that can be used to validate—with higher certainty and greater sensitivity—information obtained from sequencing and/or previous publications as noted in databases such as the Sanger Catalog of Somatic Mutations in Cancer (COSMIC).

References

1. Jurinke, C., *et al.* MALDI-TOF Mass Spectrometry: A Versatile Tool for High-Performance DNA Analysis. *Molecular Biotechnology* **26**, 147-163 (2004).
2. Yang, H., *et al.* Sensitive detection of human papillomavirus in cervical, head/neck, and schistosomiasis-associated bladder malignancies. *PNAS*. **102** (21) 7683-7688. (2005).
3. Thomas, R.K. *et al.* High-throughput oncogene mutation profiling in human cancer. *Nature Genetics*. **39**, 283-294 (2007).
4. MacConaill, L. *et al.* Profiling Critical Cancer Gene Mutations in Clinical Tumor Samples, *PLoS ONE*, (2009) 4(11) 1-7.
5. Pearce, M., *et al.* Mutation profiling in tumor samples using the Sequenom OncoCarta™ Panel. *Nature Methods*, June 2009, vii.
6. Vivante, A., *et al.* (2007) High-throughput, sensitive and quantitative assay for the detection of BCR-ABL kinase domain mutations. *Leukemia* **21**, 1318–1321.
7. Turakulov, R. *et al.* Ultrasensitive Determination of Absolute mRNA Amounts at Attomole Levels of Nearly Identical Plant Genes with High-Throughput Mass Spectrometry (MassARRAY). *Plant Cell Physiology*, 48(9): 1379–1384 (2007).
8. Mefford, H.C., *et al.* A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res.* 19: 1579-1585 (2009).
9. Rousseeuw, P.J, and Leroy, A.M. Robust Regression and Outlier Detection. Wiley, New York (1987).

The preceding references were specifically cited in this application note. For a complete list of over 500 citations, visit www.sequenom.com.

OncoCarta and TYPER are trademarks of Sequenom. Sequenom, Inc., SpectroCHIP and MassARRAY are registered trademarks of Sequenom. All products from Sequenom are for Research Use Only, not for use in Diagnostic Procedures.