「多標的陣列平台基因診斷試劑－查驗登記審查指引：（以去氧核糖核酸標誌、變異及基因表現類多標的陣列平台診斷測試為例）」

Guidance for Evaluation and Registration: Array-based Multiplex Diagnostic Kit for Genetic Tests and Gene-Expression Tests Based on DNA Markers, Mutations and Expression Patterns

第一條、目的 本指引的目的在於提供多標的陣列平台診斷試劑產品，申請上市所需之查驗登記及審查的指引。 一、以陣列(Array)為基礎的測試，例如：寡核苷酸(oligonucleotide)、cDNA、蛋白質及組織陣列都屬於多標的平行檢測的一種。 二、對於多標的平行檢測陣列平台診斷試劑查驗登記所需提供文件的建議能運用在以陣列為基礎的測試以及其他型式的多標的平行檢測。雖然本指引是以核苷酸陣列的測試為主要的考量，但相同的原則也可應用到以蛋白質或組織陣列的測試方法。	Article 1 Purpose This document is intended to provide guidance on preparing and reviewing premarket approval (PMA) submissions for array-based multiplex tests. 1. Array-based tests, such as oligonucleotide, cDNA, protein and tissue arrays, are a subset of multiplex tests. 2. The following recommendations for elements of an array-based multiplex test submission apply to array based tests as well as other types of multiplex tests. This guidance primarily considers nucleic acid based analysis, but many of the principles apply to protein and tissue arrays as well.
第二條、建立有效率的審查法規 一、由於這些技術發展出來的新興測試方法能用於疾病的診斷、預測罹患疾病的可能以及其他應用如協助使用藥物之選擇等，對醫療照護之發展具有潛在的運用價值。 二、衛生署期望與相關研發廠商合作，建立明確的法規及標準，以促進多標的平行檢測陣列平台診斷試劑產品早日上市。	Article 2 Establish the least burdensome regulations 1. Department of Health (DOH) recognizes that new tests based on this technology have the potential to enhance medical care by refining patient diagnosis for disease, disease susceptibility, and drug selection among other potential applications. 2. DOH is eager to work together with manufacturers to establish the regulations for facilitating the products of array-based multiplex tests into the marketplace.
第三條、「基因型」及「基因表現」多標的平行檢測陣列平台診斷試劑的差異 一、多標的平行檢測陣列平台診斷試劑依其功能可區分為檢測「基因型(Genetic)」及檢測「基因表現(Gene expression)」兩大類。 二、「基因型」測試：無論是生殖細胞或體細胞，其 DNA 的變化是	Article 3 The differences between ”Genetic” and ”Gene expression” array-based multiplex tests 1. In general, the array-based multiplex tests can be classed into “Genetic” tests and “Gene expression” tests by their function. 2. “Genetic” tests: DNA differences are fixed, whether germinal or somatic.

固定的。 (一)、「基因型」測試的結果一般可區分為 1、二極化(dichotomous)：有或無 2、三極化(trichotomous)：同合子 A (homozygous A)、異合子 (heterozygous)、同合子 B (homozygous B) 3、分類化(categorized)，如：單套型(haplotypes)等。 (二)、「基因型」測試的申請廠商應清楚明確的說明試驗結果所得差異現象。對於 DNA 陣列測試方法應該謹慎設計，並具有高度的再現性，其功能表現亦應明確建立。臨床研究應該說明在研究族群中的疾病盛行率。 三、「基因表現」測試：是測量基因表現改變的試驗。 (一)、相較於 DNA 的變化，基因表現的變化是會受到許多不同因素的影響。這些因素包括簡單的個體與個體間的差異、檢體採集的時間點，以及某些醫療處置對組織產生的特殊影響。 (二)、「基因表現」測試方法，在診斷、預測及篩檢，可能因這些因素的交互影響產生測試的結果有很大的不同，導致日後解釋的困難。廠商應該謹慎考慮陣列器材硬體的設計策略、品質管制(QC)、結果的再現性及數據的判讀與解釋。	(1) Results from these tests can generally be described as A. dichotomous (either present or not present), B. trichotomous (homozygous A, heterozygous, homozygous B), or C. categorized (e.g., haplotypes). (2) Interpretation of tests designed to measure these types of differences will be, in most cases, straightforward. DNA array tests nevertheless should be carefully designed and highly reproducible, and have well established performance. Clinical studies should account for disease prevalences in the populations studied. 3. “Gene expression” Tests： measuring expression changes: (1) In contrast to DNA changes, they can be responses to a variety of factors. These may include simple individual-to-individual differences, time of day, and specific effect of a therapeutic treatment on a tissue. (2) Results can vary markedly as a result of these factors. Tests to diagnose, predict, or select based on expression patterns may consequently be difficult to interpret. Sponsors of these tests should consider array physical design strategies, quality control (QC), reproducibility and readout/ interpretation.
第四條、多標的平行檢測陣列平台診斷試劑查驗登記申請所需文件： 一、當產品運用多標的平行檢測陣列平台技術，無論是檢測「基因型」或「基因表現」的差異，必須能在資料中對下述問題加以說明。 本署將針對個別產品的特性，以個案方式處理，必要時得要求不同的資料以及統計分析方法。 二、申請資料如下(但不限)： (一)、預期檢測的標的(Intended Use) (如：測定細胞色素 cytochrome P450 酵素的對偶基因(alleles)，	Article 4 Documents for the preparation of the array-based multiplex test application 1. The following are areas that we believe should be addressed in the preparation of a submission for array-based multiplex products, whether the device measures genetic or expression differences. DOH will review the application case by case and may request different types of data and statistical analyses. 2. The information requested depends on the (1) Intended use (for example, to detect cytochrome P450 enzyme alleles), (2) indications for use (for example, predictive or prognostic for

(二)、宣稱使用的適應症(如：對疾病狀態、反應或敏感度的預測或預後)， (三)、使用的方法(如：聚合酶鏈鎖反應)， (四)、結果的判讀(如：對各種對偶基因的陽性反應)， (五)、功能的表現(如：分析方法確效、品質管制及試驗的限制)， (六)、臨床正確性(validity) (如：偽陽性或偽陰性)， (七)、臨床判讀(如：效益及風險)， (八)、廠商的宣稱用途(如：有效性)：	disease, response, or sensitivity), (3) methodology (for example, polymerase chain reaction), (4) technical interpretation of results (for example, positive for variant alleles), (5) performance characteristics (for example, analytical validity, quality control and assay limitations), (6) clinical validity (for example, false positives and negatives), (7) clinical interpretation (for example, benefits and risks) and (8) claims made by the manufacturer (for example, effectiveness).
第五條、說明事項的建議 一、申請產品之預期檢測的標的應能詳細描述該測試所能測量的項目，為何進行測量的理由以及該測試所針對的特定族群以及採用的理由。 二、某些測試也許有多種預期檢測的標的。因此對每一項預期檢測的標的應該分別提出申請，並提供各自獨立且不同的研究資料結果來支持。對不同的測試方式或多項的預期檢測的標的提出申請時，應與衛生署及所屬的相關單位諮詢與討論。	Article 5 Recommendations for addressing these issues follow. 1. The intended use should specify what the test is intended to measure, why it is measured, and should specify populations to which the test is targeted, where appropriate. 2. Some tests may have multiple intended uses. DOH recommends a separate application for each intended use that requires unique and separate supporting studies. You should consult the appropriate review divisions in DOH for advice on submitting tests with multiple intended uses.
第六條、分析方法確效 一、設計與製造 (一)、多標的平行檢測陣列平台診斷試劑為醫療器材，其設計、製造與管制應符合「醫療器材優良製造規範」 (Quality System)的規定。 (二)、詳細說明多標的平行檢測陣列平台診斷試劑所使用的組成：如 1. 產品設計、 2. 使用的內部對照組、 3. 寡核苷酸、 4. 引子、 5. 探針、	Article 6 Analytical Validation 1. Design and Manufacturing： (1) Array-based multiplex tests are medical devices, their product design manufacture and control should fulfill the “Good Manufacturer Practice for Medical devices” Quality System (QS). (2) Specifically, the following elements of array-based multiplex products should be well characterized: A. design, B. internal controls used, C. oligonucleotides, D. primers, E. probes, or F. other capture elements,

6. 其他捕捉物質、 7. 製造陣列的管制條件(包括清洗的過程、乾燥條件如溫度及時間)、 8. 將標的物與基質結合的方法、 9. 陣列或其他固定空間平台的組成及排列方式、 10. 標誌或標的物的專一性及平台的安定性。 (三)、廠商在提出查驗登記申請時，應有分析方法資料來顯示該器材能正確的運作以及在特定的條件下其功能的表現是可信賴的，資料內容包括： 1. 檢體或樣品(對每個宣稱的組織間質):對於樣本用在陣列中進行雜合(hybridized) 等需提供包括鑑別、配製、適當的的接受條件、決定標記結合程度的測量方法、探針的長度等資料。同時也包括檢體的採集、儲存及處理的條件。 2. 測試組件 (Assay components): 包括：緩衝液、酵素、訊號檢測系統(包括螢光染料、冷光試劑、其他檢測試劑、器材或軟體)。 3. 標準品及(或)校正品 (Controls and/or calibrators):用作內部或外部的陽性及陰性標準品 二、對特定的功能表現進行確效試驗：分析實驗室研究 對於每項預期檢測的標的，廠商應提供每個標靶、特徵、標誌或突變標記的表現功能經由分析實驗室的研究結果來說明，包括： 1. 分析靈敏度: 能正確檢測陽性樣品的能力 2. 再現性 3. 閾值、參考範圍、或醫學判定點的確效 4. 測試範圍 5. 樣品過量或過少時的影響，未能產生再現性可接受結果時，應探討樣本濃度，以產生再現性。	G. conditions for producing arrays, including washing procedure and drying conditions (e.g., temperature, length of time), H. methods used to attach the target material to the matrix, I. composition and spatial layout of arrays or other spatially fixed platforms, J. specificity for markers or targets, and stability of the platform. (3) We recommend that submissions include analytical data that demonstrate that the device performs accurately and reliably under given conditions; this may include: A. Specimen/sample (for each claimed matrix): identity, preparation, acceptance criteria where applicable, and methods for determining label incorporation, probe length, and so forth, for samples that will be hybridized to the array. Also, include specimen collection, storage, and handling conditions. B. Assay components: including buffers, enzymes, signal detection systems such as fluorescent dyes, chemiluminescent reagents, other signaling reagents, instruments, and software. C. Controls and/or calibrators: negative and positive controls, characterized as internal or external. 2. Validation of Specific Performance Characteristics: Analytical Laboratory Studies Sponsor Should describe the following performance characteristics for each target, pattern, marker or mutation claimed in the intended use statement: A. Assay sensitivity: ability to accurately identify positive samples. B. Reproducibility: C. Validation of cut-off, reference range, or medical decision point. D. Assay range. E. Effect of excess sample and limiting sample. Investigate the sample concentrations and conditions that reproducibly yield acceptable results F. Assay specificity and interfering substances (endogenous and

6. 測試結果的分析專一性及可能的干擾物質(包括內生性及外源性)	exogenous).
第七條、陣列及數據之處理 一、陣列及數據之資料： 1. 同時對多個標的進行檢測或區分方法的最佳化條件，如：雜合條件、反應物的濃度以及專一性的控制。 2. 檢體間之交叉污染。 3. 資料處理的計算方法。 4. 陣列的限制因素：包括雜合反應的飽和程度。 二、提供下述器材的確效資料： 1. 特性：使用在測試過程的器材特性，包括器材如何判讀數值或對測試結果的解釋，例如：落點、大小、濃度、體積、小量樣品乾涸所產生的問題、小量容積的反應效果以及對試驗結果的影響。 2. 器材校正: 對器材校正加以描述。 3. 不確定性: 說明使用硬體時對結果判讀產生不確定因素的原因及估計其不確定性。	Article 7 Array and data processing 1. Sponsor should describe： (1)Optimization of multiple simultaneous target detection/differentiation, for example, hybridization conditions, concentration of reactants, control of specificity. (2)Potential for sample carryover. (3)Computational methods for data processing. (4)Limiting factors of the array, including saturation level of hybridization. 2. Validation of instrumentation Submission include the following: (1) Characterization: Characterize instruments used in the assay, including how the instrument assigns values to or interprets assay variables such as feature location, size, concentration, volume, drying of small samples, effect on small volume reactions and its impact on test results. (2) Calibration: Describe instrument calibration. (3) Uncertainties: Describe sources and estimates of uncertainties in results introduced by hardware components.
第八條、用臨床檢體進行比較研究 當有適當的評估方法可用來與該器材之功能進行比較研究時，應提供下列的資料來提出申請： 一、與經核准之醫療器材進行比較試驗的結果，經常以其結果的一致性百分比表示。 二、與經認可的標準方法或臨床診斷結果進行比較試驗的結果，經常以專一性及靈敏度表示。 三、比較結果的差異性及其解析方法應該採用正確及適當的統計方法來執行。	Article 8 Comparison studies using clinical samples Where comparison studies are appropriate to establish performance of a device, the following items could be used to support submissions: 1. Comparison to another device: Results of comparison studies with another well-characterized or predicate device; usually reported as percent agreement. 2. Comparison to a Reference Method: Results of comparison studies to a validated reference method or clinical diagnosis; usually reported as sensitivity and specificity.

四、確認在分析方法或技術上會產生偽陽性及偽陰性之結果，並預估測定失敗的比例。 五、以量化之測定方式來評估試驗結果 在與真據或標準方法的比較中評估產生的隨機及系統誤差。	3. Resolution of Comparison Discrepancies: Results of discrepancy testing should be reported; resolution should be performed only using unbiased statistical techniques. 4. Identification of analytical/technical false positive or false negative results, estimates of expected assay failure rates. 5. Evaluation of tests employing quantitative measurement techniques: evaluation of random and systematic error in comparison to the predicate or reference method.
第九條、與已認定之診斷方法進行臨床評估之比較研究 當需要執行臨床試驗以建立多標的平行檢測陣列平台器材類產品的安全性及有效性時，需提供下述資料 一、支持預期檢測標的之臨床資料應該提供適當的臨床資料來支持每一項預期檢測標的。某些情形下，可以直接引用專業領域對此預期檢測標的之敘述或指引文件(guideline)。 當無充分之臨床試驗資料或臨床資料不足（如：有人種差異、感受性病原體差異時），需在國內執行臨床試驗。建議廠商先與衛生署或所屬相關單位進行諮詢後，再提出申請。 二、衛生署建議廠商在產品申請時應提供下述資料來支持臨床確效： 1. 受試者同意書(Informed Consent)及符合人體試驗委員會(IRB)相關規定：用在進行器材確效的臨床樣本應符合醫療法、藥事法、優良臨床試驗規範及衛署醫字第０九一００一二五０八號公告之【研究用人體檢體採集與使用注意事項】等相關規定。 2. 「臨床真據」“Clinical truth”： 廠商應對用在評估所申請之器材的臨床功能，所採用的「臨床真據」加以說明。 3. 臨床資料 (Clinical data)：應對基因表現的模式、基因型/基因表現的相關性等資料進行確效，如採用其他檢測系統(定量RT-PCR)進行確證。對每一項預期檢測的標的 (包括在各項	Article 9 Clinical Evaluation Studies Comparing Test Performance to Accepted Diagnostic Procedure(s) Where clinical studies are necessary to establish safety and effectiveness of a array-based multiplex device, you should address the following points: 1. Clinical Data to Support intended Use. You should provide appropriate clinical data to support each intended use. In some cases, it may be appropriate to include a direct reference to a professional statement or guideline in the intended use statement. If the clinical data is lacking or insufficiency (e.g., ethnic factors or sensitive pathogen has regional differences), local clinical trial(s) are required. We encourage sponsors to consult with DOH or DOH assigned third-party (such as Center for Drugs/Devices Evaluation [CDE]) to determine the suitability of reference to such statements or guidelines. 2. Clinical Validation: DOH recommends that the following items be addressed in the submission to support clinical validation: (1) Informed Consent and Investigational Research Board (IRB) Requirements: Samples that are used in the clinical portion of the device validation must be obtained in conformance with DOH requirements (eg. Law of Medical Affairs; Law of Pharmaceutical Affairs, Good Clinical Practice and Office Announcement of No. 0910012508 from Medical affairs [Guideline for Collection and Utilization of Human Samples]).

宣稱用途的陣列試驗中所採用的臨床樣品)所用檢體的適當數量都應有統計學的基礎。當定義試驗族群時，則應提供下述資料： (1) 自正常族群中抽取的樣本數，並摘要樣本的人口數據特性。 (2) 在每一種疾病、不同情況、病原、基因型或組別內採用的檢體數，並個別摘要其人口數據特性。 4. 標準品的範圍(Reference ranges)：在可能的情況下，正確計算標準品的範圍。 5. 統計方法 (Statistical method)：應說明用於計算的統計方法。精確性的測量方法，如：信賴區間等，應該加以描述及提供數據。 6. 文獻 (Literature)：在以陣列為基礎的測試系統中，某些標誌、突變或表現模式，可能有足夠的文獻資料作為新的測試方法建立臨床有效性的依據。 如果申請廠商希望引用文獻來支持臨床有效性，則應提供對相關的已發表及未發表的文獻資料作成摘要說明，加上已發表文獻中對所申請器材的臨床資料。 (1) 當希望提供文獻作為銜接實驗室分析到臨床功能表現的資料說明時，該文件資料所用的技術應與新申請產品的技術相同，同時所研究的病人族群也相似。 (2) 文件是否得代替或補充臨床效能的評估試驗，廠商可以尋求衛生署及所屬單位的諮詢。	(2) “Clinical truth”: Define clinical truth as it will be used in evaluating the clinical performance of the device. (3) Clinical data: Validate expression patterns, genotype/phenotype correlations, and so on, on a statistically adequate number of specimens for each intended use, including clinical samples for all matrices claimed in the intended use statement; verify with a second detection system (e.g. quantitative RT-PCR) if applicable. When defining the populations used, submissions should include the following information: A. Number of samples from the normal population with samples summarized according to appropriate demographic characteristics. B. Number of specimens included in each disease, condition, pathogen, genotype, or group summarized according to appropriate demographic characteristics. (4) Reference ranges: Calculate reference ranges when appropriate. (5) Statistical method: Describe statistical methods used for calculations. Measures of precision, e.g., confidence intervals, should be described and presented. (6) Literature: For some markers, mutations, or patterns in an array-based test system, there may be a sufficient literature base to establish clinical validity with the new test. If a sponsor intends to use literature to support clinical validity, include a summary of available published and unpublished information and/or published clinical data pertinent to the device A. When literature is intended to support bridging from analytical to clinical performance, the literature should identify the same technology as the new test and a similar patient population. B. We recommend that you consult DOH or CDE to determine the suitability of literature to supplement or substitute for clinical performance studies..
第十條、診斷器材的臨床有效性	Article 10 Clinical Effectiveness of the Device

一、新標誌 對於應用多標的平行檢測陣列平台診斷試劑來評估新的標誌、突變、模式或其他檢測之結果應符合本署有關臨床有效性之規定。 三、已被認可之標誌 當分析效能已於宣稱的相關樣品組織間質中獲得確認，，申請廠商可以提出醫學文獻作為該項標誌或突變點測試有效性的證據 1. 如果廠商提供嚴謹可信的文獻作為臨床有效性的證明時，應該提供所有相關文獻的影本資料，以及說明利用這些文獻來取代臨床試驗的理由及其正當性。 2. 同時也應證明所申請的新器材與文獻中所使用器材的效能相當，以確保該類資料外推應用到宣稱產品時，仍能正確無誤。	1. New markers Evaluations of new markers, mutations, patterns, or other outputs of multiplex tests should meet the DOH standard for clinical effectiveness for their intended use. 2. Established Markers When analytical performance is validated in the specimen matrix claimed, the sponsor may use the medical literature as evidence of the effectiveness of the marker or mutation. (1) If a sponsor wants to use peer-reviewed literature to support effectiveness, you should furnish copies of all relevant articles and provide a justification for the use of the literature in place of clinical studies. (2) The sponsor should establish comparability between the new device and the device used in the published literature in order to ensure that the data can be confidently extrapolated.

附錄一：設計及評估臨床試驗的一般考量	Appendix I: General considerations for planning and evaluating clinical studies.
下列的一般建議可以作為設計及評估臨床試驗的考量。衛生署建議廠商可以向醫藥品查驗中心來進一步諮詢執行臨床試驗的策略及方法。 一、詳細說明體外試驗臨床評估計劃書。明確的定義試驗族群以及納入(inclusion)及排除(exclusion)條件和所選定的臨床指標。如果引用文獻，則研究族群、納入/排除條件及臨床指標應該在文獻中有明確的解釋，同時能適當的反應為何該器材能應用在臨床。同時該研究族群及臨床指標應該與廠商所預期檢測的標的及用途相符。 二、選擇適當的試驗中心，以進行所預期檢測標的及用途之臨床試驗。當試驗族群取樣的偏差可能影響到臨床試驗的結果時，應在計劃書中詳細說明。 三、在執行臨床試驗前，所有外部評估中心應建立一致的臨床試驗計劃書，並在資料蒐集期間應定期的進行監督與查核。 四、在執行臨床試驗前應先決定試驗的樣本數。對每一標誌、突變或表現模式，樣本數均應有足夠的統計檢定力來檢測出臨床上有意義的差異。如果能提供的樣本數量過小時(如：該疾病或生理狀態的盛行率很低；或是標誌或突變的機率很低時)衛生署可能視情況要求其他資料來源。 五、說明在臨床試驗中選擇或排除病人的取樣方法。如果需要使用到庫存的樣本或進行回溯性實驗設計時，必須提供充足的理由說明這些樣本族群與病人族群間的關係。 六、如果臨床和統計上有此必要，應同時根據個別及整合各試驗主持人/中心兩種方式，分析試驗數據。對於具遺傳性的標誌或突變，如果整合分析數據即可接受，則性別、種族或人種的人口數據特性在各試驗中心均應該相似。 七、儘可能將基因型資料以適當的 N x N 表格呈現(如：用 3 X 3 表	The following are general recommendations that may be used when planning and evaluating clinical studies. We recommend that you consult with CDE reviewers' divisions to determine the most appropriate strategies for clinical studies. 1. Describe all protocols for external evaluation studies. Clearly define the study population and inclusion and exclusion criteria and the chosen clinical endpoint. If literature is to be used, the study population, inclusion/exclusion criteria, and endpoints should be clearly explained in the publication and be reflective of how the device will be used in practice. The study populations and endpoints should correspond to the intended use and claims of the manufacturer. 2. Use investigational sites appropriate to the intended use and claims being sought. Efforts to define population sampling bias should be clearly outlined when this issue may affect performance. 3. Establish uniform protocols for all external evaluation sites prior to study and follow them consistently throughout the course of data collection. 4. Determine sample size prior to beginning the clinical study. The sample size should have sufficient statistical power to detect differences of clinical importance for each marker, mutation, or pattern. DOH will consider alternate data sets incases with a small available sample size, for example, a disease or condition having a low prevalence or with markers or mutations of very low frequency. 5. Describe the sampling method used in the selection and exclusion of patients. If it is necessary to use archived specimens or a retrospective design, provide adequate justification for why the sampled population is relevant to your patient population. 6. Analyze test data both by separate investigator/site and pooled over investigators, if statistically and clinically justified. For heritable markers and mutations, gender and race or ethnicity demographics

格說明同合子野生型，異合子及同合子突變型)。 八、所提供的資料應在該器材適用的族群中具有代表性以支持其所預期檢測的標的。如果因為人種差異而在標誌/突變點產生不同，則須於試驗中納入不同的人種。 九、儘可能提供因為疾病或其他狀態而導致測試時會出現偽陽性或偽陰性的個體樣本(經由鑑別診斷所判定的結果)。 十、應包括所有的病人與檢體。作適當的資料稽核及驗證。如果病人或樣本經收納後，其結果於其後被排除，則應提出具體說明理由。 十一、執行臨床試驗時應有正確的品質管制。並說明執行品質管制所用的材料及方法。 十二、說明如何決定閾值(常作為區別陽性與陰性樣本的判斷值或是醫學判定的極值)的方法。描述個別標誌/突變點在閾值決定後其性能表現的特性。在描述如何決定個別的閾值時應包括使用的統計方法(如：ROC 曲線法(Receiver Operating Characteristic Curve))。 十三、無論適用在藥品研發或診斷目的的微陣列研究，其所提供的材料來源、資料種類和相關遺傳資訊應符合「微陣列實驗基本資訊規範(Minimum Information About a Microarray Experiment” MIAME)」(www.mged.org/miame)	markers and mutations, gender and race or ethnicity demographics should be similar between sites if data pooling is otherwise appropriate. 7. Display genotype data in the appropriate Nx N table (e.g., 3 X 3 for homozygous wildtype, heterozygous and homozygous mutation) where applicable. 8. Support the intended use claim for the device with data that are representative of the population for whom the device is intended. Include a diversity of ethnic groups if the marker/mutation varies according to ethnicity. 9. Include samples from individuals with diseases or conditions that may cause false positive or false negative results with the device (i.e., within the differential diagnosis), if appropriate. 10. Account for all patients and samples. Perform appropriate data audits and verification before submitting to DOH. Give specific reasons for excluding any patient or result after enrollment. 11. Perform studies using appropriate methods for quality control. Describe the materials and methods used to assess quality control. 12. Describe how the cut-off point (often the distinction between positive and negative, or the medical decision limit) was determined, if appropriate. Describe the performance characteristics the cut-off identifies for each marker/mutation. The description of how each cut-off was determined should include the statistical method used (e.g., receiver operating characteristic curve). 13. The “Minimum Information About a Microarray Experiment” (MIAME) guidelines (see www.mged.org/miame for more information) describe many of the sources and types of data and information that should be available for most types of microarray studies, whether they are used to support drug development or diagnostic device submissions.

附錄二：分析陣列數據之統計之考量	Appendix II: Statistical Considerations for Analyzing Array Data
一、檢測基因表現的陣列晶片(expression arrays)：採用適當的統計方法鑑別不同的族群(如：正常組、疾病組)，其中包括監督分析(如：鑑別分析、多項迴歸分析及支援向量模式(support vector machines) (一)、同時亦可考慮發現新群組的未監督分析方法，如：集群分析(Cluster analysis)；型態識別(Pattern Recognition)；自組織映射圖網路(Self-Organizing Maps)及因素分析(Factor Analysis))以作為建立診斷分類的基礎。這些可能十分複雜的方法，通常只用來設定閾值； (二)、對於測試方法的表現評估，則以簡單的統計方法，如：敏感度、專一性、一致性百分比、陽性及陰性預測值等，進行評估。 (三)、評估閾值表現所用的資料來源應與用在建立閾值的資料來源不同，否則其表現的評估會被高估。處理的方法之一，可用統計方法來校正此偏差(如：the leave-one-out method；「摺刀」(jackknife)方法；拔靴法(Bootstrap)。ROC 曲線法(Receiver operating characteristic curves)有助於評估產品表現，但在應用上仍需預先選擇適當的閾值。 二、統計分析應該考慮下列不獨立的狀況，如：同一片晶片內不同重複；執行內之樣本；不同日內多次執行之數據的相關性。例如：變異數成分分析(variance component analysis)可解決數據相關性之問題。多重之單一核苷酸多型性(single-nucleotide polymorphism, SNP)分析方法，或以 DNA 為主的多標的平行檢測陣列平台診斷性試劑所產生的多變數的測量值，將使比較不同試劑更為複雜。對於多變數測量值，提供一個測量值的摘要統計量有助於結果之比較。 三、方法的比較(Method Comparisons)：在無已知真據下所作之試驗比較只是對一致性的比較，會有下列限制： (一)、因為真值(實際診斷或定量測量)未知，因此方法的比較只能建	1. Expression arrays: Appropriate statistical analyses for discriminating subjects into groups (e.g., normal, diseased) include supervised analyses (e.g., discriminant analysis, multinomial regression, support vector machines). (1) Unsupervised analyses that allow for discovery of new groups would also be considered (e.g., cluster analysis, pattern recognition, self organizing maps, factor analysis) as a basis for building diagnostic categories. While such analyses can be complex, they are generally used only to establish the cut-offs for discriminating between pre-specified groups; (2) actual performance of tests could be evaluated with simple statistical analyses of sensitivity, specificity, percent agreement, positive and negative predicted values, among others. (3) Evaluation of the performance of cut-offs should be based on a dataset that is independent of the dataset used to establish the cut-offs, otherwise the performance will tend to be overstated. Alternatively, statistical methods can be used to correct for this bias (e.g., the leave-one-out method, the jackknife, or the bootstrap). Receiver operating characteristic curves are useful for evaluating performance, but cut-offs still need to be chosen to apply the test in practice. (2) The statistical analysis should account for lack of independence due to, among others, correlation of replicates within chips, samples within runs, runs within days. For example, a variance component analysis could be used to account for correlation. Multivariate measurements by methods such as multiple SNP analysis or multiplex DNA-based tests, complicate comparison of tests. For multivariate measurements, a summary of the measurements could be helpful in making comparisons. 3. Method Comparisons: Comparisons of tests without a measure of truth are comparisons of agreement and have the following

立相等性而非優越性。 (二)、資料的一致並不代表正確，因為可能兩種方法所得的值雖然不正確但仍一致。 (三)、對於診斷的結果，一致性的程度是與疾病的盛行率有關，因為一致性判斷的依據是在最後真正診斷結果為陽性或陰性。對於定量測量的結果，一致性的程度是與測量值的大小有關。當某一變數的一致性分歧時，統計分析應對此變數執行分層分析。 (四)、對於診斷結果，相對的靈敏度及專一性以及差異性的解決方法可能誤導且不適於作為是否可以核准的主要評估標準。因此對於陽性及陰性結果一致性的比率應該提供給衛生主管單位。 四、目前新的統計方法已發展出在真值未知之前即可比較不同測試的方法。這些方法有助於估算未知的真值及比較兩種方法間的優越性。然而這些方法需要強烈假設如試驗結果間的相關性、未知真值的分布(如:盛行率)、以及標準或真據方法的功能表現(如：敏感度和專一性)。這一類的假設必需加以證實。 五、對於沒有真正標準的定量方法比較時，繪圖對於將一致性的誤差(error of agreement)分為系統性及隨機性的誤差(systematic and random error)的分析判斷是非常有幫助的。在以實驗方法的測量結果作為 Y 軸，與相對的控制方法(標準或真據)的結果作為 X 軸所得的散佈圖(Scatter plot)中，系統性誤差可以由比較相等線(y=x)與散佈點來評估。在相等線上表示未出現系統性誤差。隨機性誤差(Random error)可由資料的散佈來評估，資料的分佈表示方法間的變異。利用 Bland-Altman 散佈圖，將兩種方法的配對測量值以其差值及平均為兩軸，對於測量範圍內系統性及隨機性誤差趨向的檢測特別有用。 六、使用虛無假設(null hypothesis)檢定：以正式的統計分析來檢定實驗方法與控制方法的相等性(equivalence)。例如：兩種測試方	limitations: (1) Because the true value (the diagnosis or the quantity being measured) is unknown, comparison of the methods can only establish equivalence, not superiority of one method over another. (2) Agreement is not a measure of correctness because both methods could agree on an incorrect value. (3) For diagnosis, level of agreement usually depends on prevalence because it depends on whether the true diagnosis is positive or negative. For quantitative measurement, level of agreement often depends on the magnitude of the measurement. When agreement is heterogeneous over a variable, its statistical analysis should be stratified by that variable. (4) For diagnosis, relative sensitivity and specificity and discrepant resolution can be very misleading and are not appropriate for primary evaluation of approvability. DOH recommends reporting of positive and negative percent agreement. 4. Recently, statistical methods have been developed that allow comparison of methods with the unknown true value being measured. These methods might be useful for obtaining estimates of performance with respect to the unknown true value and for establishing superiority of one method over another. However, these methods often make strong assumptions about the correlation between test results, the distribution of the unknown true values (e.g., the prevalence), and the performance of the reference or predicate test (e.g., its sensitivity and specificity). Such assumptions need to be justified. 5. For quantitative method comparison data without a truth standard, plots are very useful for decomposing the error of agreement into systematic and random error. In a scatter plot of the experimental method measurement (y) versus the corresponding control (reference or predicate) method measurement (x), systematic error can be evaluated by comparing the scatter with the identity line

法是否定義為相等，可由其配對測量值(paired measurements)做線性回歸後的斜率(b)是否接近 1 來判斷。根據前述的定義，一個能有效證明相等性的方式是設定虛無假設為斜率(b)與 1 的距離超過 d；d 是事先指定之最小臨床上有意義的差距。如果虛無假設被推翻，則顯示兩種方法已證明具相等性。另一個常用但不適用的方式是設定虛無假設為 b 等於 1。不適用的原因是當沒有足夠的證據去推翻虛無假設時，並不代表有足夠的證據去接受該假設。事實上，當樣本數太小時通常會發現無法推翻虛無假設。 七、其他考量點：實驗方法與控制方法的一致性薄弱時，有可能只因為來自控制方法的再現性不佳，即使實驗方法的測量結果與真值非常吻合。在相同情形下執行重複的測量可能有助於發現這種問題。這種測量的變異性或誤差，在以標準線性回歸比較兩種方法時，也會造成斜率的低估。可考慮測量誤差的其他回歸方法，如：戴明回歸分析方法(Deming regression)及 Passing-Bablok 二氏回歸分析方法(Passing-Bablok regression)。	(y=x), which indicates no systematic error on average. Random error is assessed by the spread of the scatter, i.e., the variability between the methods. A Bland-Altman scatter plot of the difference between paired measurements from the two methods versus their average is especially useful for detecting trends in systematic and random errors over the measurement range. 6. Use of null hypothesis testing: Formal statistical analyses test equivalence of the experimental method with the control method. For example, two methods could be defined as equivalent based on the slope b, from a linear regression of paired measurements, being close to one. For this definition, a valid approach tests the null hypothesis that the b is more than d units away from one, where d is pre-specified to be the smallest clinically meaningful difference. Rejection of this null hypothesis then implies that equivalence has been demonstrated. A common, but invalid, approach tests the null hypothesis that the b equals one. This approach is invalid because insufficient evidence to reject the null hypothesis does not imply sufficient evidence to accept it. In fact, a sufficiently small sample size can be chosen to guarantee that the null hypothesis will not be rejected. 7. Other considerations: Poor agreement between the experimental and control methods might be simply due to poor repeatability of the control method, even if the experimental method measures the true value perfectly. Duplication of measurements under the same conditions may be needed to identify this problem. This variability or error in the measurements also biases downward the estimate of the slope in a standard linear regression comparing two methods. Alternative regression methods that account for measurement error include Deming regression and Passing-Bablok regression.