

# 知識天地

## 從DNA差異閒談個人化醫學

施柔合(小姐)<sup>1,2</sup>、谷德倫(先生)<sup>1,3</sup>、周玉山(副研究員)<sup>1,2</sup>

<sup>1</sup>本院生物醫學科學研究所

<sup>2</sup>國立臺灣大學基因體與系統生物學學位學程

<sup>3</sup>國立陽明大學微生物及免疫學研究所

從第二次世界大戰結束到1970年間，許多遺傳生物學家發現到突變對於疾病以及存活率的關聯性，但因受限於技術層面而未能有所突破。但是到了1970年中晚期，隨著基因重組技術(Recombinant DNA technology)、聚合酶鏈鎖反應(Polymerase chain reaction, 簡稱PCR)、及核酸定序(DNA sequencing)等技術的發明及成熟，將人類全基因體約 $3.3 \times 10^9$ 核酸序列完全解碼，並探討基因突變與人類疾病致病的關聯性是解決人類疾病負擔的終南捷徑。因此到了1990年，生物學家開始著手人類全基因體定序計畫(Human Genome Project, HGP)，並在2001年宣布完成人類全基因體序列草圖。

然而人類基因體序列完全解碼只是取得一份標準的人類基因體核酸序列藍圖(standard blueprint)，利用蒐集疾病家族尋找突變(mutation)基因並研究突變造成疾病的致病分子機轉，可進一步研發疾病篩檢、預防和治療的技術及藥物，減輕及消彌人類對疾病的負擔。但是家族性的遺傳疾病通常是單一基因突變，其突變造成的病症通常比較嚴重，罹病的人也較少，因此歸類於罕見疾病，例如杭亭頓舞蹈症、黏多醣症、玻璃娃娃等。OMIN (Online Mendelian Inheritance in Man)是收集這些罕見疾病和其基因突變最常用的資料庫。

但是大多數人的疾病是「多個基因突變的組合」加上環境因素的配合所造成；例如氣喘是常見的人類疾病，而氣喘發作除了遺傳因素外，也因為過度暴露於過敏原而發病。其他如高血壓、糖尿病、牛皮癬等皆屬於多個基因突變組合型的人類常見疾病。以高血壓為例、目前已推論出的高血壓風險基因突變超過300個，但究竟需要多少個基因突變的組合及環境因素的配合才會讓高血壓發病則尚無定論。因此這類型的基因突變會出現在「高血壓病人基因體」較「無高血壓人的基因體」為多，而導致罹患高血壓的風險。而這類型的基因突變，常出現在人類基因體中。當此常見的基因突變的出現頻率超過1%時，此基因突變改稱為「單一核苷酸多型性」(single nucleotide polymorphism, SNP)。SNP是人類基因體最常見且分佈最廣泛的變異(variation)，平均每300個鹼基對就可能有一個SNP，到2011年止，已經有52,000,000個SNP在人類基因體被偵測。由於人類遷移歷史及生活環境的不同，SNP在各人種基因體的出現頻率也不同。因此各人種長相不同、對過敏原及藥物的感受性不同、其疾病成因也可能不同。為了探討各人種基因體SNP出現頻率及其組合型態差異，「國際人類基因體單體型圖合作計畫」(International HapMap Project)，將人類基因體的變異依不同人種做編錄，而常見的SNP可製作成高密度之晶片，加速疾病風險基因突變SNP的發掘(genome-wide association studies, GWAS)。由於SNP有種族差異性，研究人類常見疾病風險基因座(risk alleles)的論文在過去5年中如雨後春筍般被大量發表，雖然這些風險基因座不見是造成常見疾病的病因，但對常見疾病的早期檢測及預防卻有重要助益。

然而人類基因體序列變異(variations)也不單單只有突變(mutation)與單一核苷酸多型性(SNP)。另一類常見人類基因體序列變異則是重複序列(repeats)和拷貝數變異(copy number variation, CNV)。常見核苷酸重複序列又稱為微衛星標竿(microsatellite marker)包含對遺傳研究貢獻良多且分布廣泛的CA重複序列(CA repeats)、造成杭亭頓舞蹈症及其它漸進性神經退化疾病的三核苷酸重複序列(NNN repeats)、和常用於親子血緣及犯罪DNA鑑定的四核苷酸重複序列(NNNN repeats)。另外整條染色體拷貝數變異則有三條21號染色體的典型唐氏症。然而「長DNA片段的重複序列」又稱為拷貝數變異(CNV)則因為技術性的困難而較晚發現。2006年數個基因體研究團隊藉由HapMap所提供的不同人種樣本分析整理出人類DNA序列中的拷貝數變異(copy number variation, CNV)和其在遺傳疾病扮演的角色。CNV是透過DNA片段缺失(deletion)、放大(amplification)、不同層級的插入(insertion)和複寫(duplication)在基因

體的許多位置得到(gain)或失去(loss)同源序列而形成多樣性的CNV。雖然CNV出現頻率遠低於SNP，但因為CNV屬於長DNA片段的重複序列，其所包含的DNA範圍可能超過10%人類基因體，更突顯CNV在基因體序列變異的重要性及對人類疾病的影響。例如恰克-馬利-杜斯氏症(Charcot-Marie-Tooth disease)有70~80%病人是肇因於17號染色體(17p12)上不正常複寫長片段DNA；有1%左右的精神疾病患者包括自閉症、精神分裂症、和其他精神疾病在染色體16p11.2、1q21.1或15q13.3會出現罕見的DNA片段缺失。

癌症是國人十大死亡原因之首，癌症的形成源自於不正常腫瘤細胞的增生，而隨後的癌細胞轉移更是癌症病人衰竭致死的主因。由正常體細胞癌化到不正常腫瘤細胞並轉移是經由不斷累積基因體變異所導致。這些後天癌化過程不斷累積的體細胞基因體變異當然也包括突變及拷貝數變異。國人常見的非小細胞肺癌(NSCLC)中表皮生長因子接受器(epidermal growth factor receptor, EGFR)的基因點突變及缺失是治療肺癌的標靶基因，其標靶藥物艾瑞莎(Iressa)可抑制因突變不斷活化的EGFR癌化訊息，有效控制肺癌生長。而乳癌病人中約有20-30%患者會有HER-2基因放大(amplification)和過度表現的現象，其標靶藥物賀癌平(Trastuzumab)可抑制HER-2不斷活化的癌化訊息，抑制乳癌生長。通常遺傳性的拷貝數變異稱為CNV(copy number variation)，但是腫瘤細胞癌化過程所累積的拷貝數變異則區別為CNA(copy number alteration)。我們實驗室利用高密度SNP晶片偵測腫瘤尤其是肝癌和肺癌全基因體變異，我們特別重視套數增加(amplification)及套數減少中的同源性缺失(homozygous deletion)區域，這些CNA區域也常常分別伴隨著致癌基因(oncogene)的活化和抑癌基因(tumor suppressor gene)的功能喪失，以發掘新穎癌症標靶基因。

由於人類基因體序列變異因種族不同而有差異，除了大家長相有異外，不同人種及個體對環境和藥物的感受性也不同。而國人常言的「體質」差異應該可用基因體變異來解釋。最顯著的例子是，非小細胞肺癌(NSCLC)中EGFR突變是治療肺癌的標靶，但EGFR肺癌基因突變卻僅好發於亞裔、不抽菸的女性肺腺癌患者。而隨著高通量核酸定序及高密度探針晶片技術越趨成熟，加速國人基因體序列變異及標靶基因的發掘是發展國人個人化醫學的重要課題。便宜的高通量核酸定序技術及生物資訊的配合，將可針對個體基因體差異發展個人化醫學來進行疾病預防、對症下藥、提升治療成效、降低疾病負擔、最後增進國人健康福祉。