RESEARCH ARTICLE

# Whole Genome Sequencing Progress in Medical Application

Marico Wagner, Daphne Benjamin
Scotland University of Life Science Institute

***Abstract:*** At present, there are still a lot of diseases pathogenesis not clear, but making use of a new generation of sequencing technology to organisms conduct whole genome sequencing, which for the pathogenesis of many diseases provide a new theoretical basis. So, Which aspects the whole genome sequencing in medical use? In this article, I got through reading the whole genome sequencing of research papers about abroad nearly five years finding that the whole genome sequencing could be widely used in genetic disease, tumor, infectious diseases, infectious epidemic, judge the individual susceptibility to disease, the diagnosis and treatment of a variety of diseases such as biological evolution. This paper reviewed the whole genome sequencing progress in medical application from a genetic disease, tumor, infectious diseases, infectious epidemic, determine the individual susceptibility and biological evolution.

***Keywords:*** Genome-wide; Whole genome sequencing; Whole genome sequencing applications

The genome refers to all the genetic material of living organisms, organelles or viruses. Whole-genome sequencing is a new generation of high-throughput DNA sequencing tools to analyze the structural differences between genomes of different organisms by means of bioinformatics, and to complete the annotation of single nucleotide polymorphism (SNP) and genome structure. Next generation sequencing (NGS) technology, combined with new pattern recognition methods and network analysis, has radically changed the understanding of basic biological mechanisms and cellular pathways. Whole genome sequencing provides a new method for gene diagnosis, making it easier to find gene mutants, which opens up a new way for gene therapy. Gene diagnosis and gene therapy are widely used in genetic diseases, tumors, infectious diseases, infectious epidemics, judging individual disease susceptibility, biological evolution and other fields. This provides a new basis for the prevention, diagnosis and treatment of many diseases, and also brings good news for many patients. This article will review the progress of whole genome sequencing in medical applications.

## 1. Genome sequencing in medical applications

The new generation of genome-wide sequencing has made genome-wide sequencing widely used in cancer, genetic diseases, infectious epidemics, infectious diseases and other diseases, as well as to determine the susceptibility of individual diseases and biological evolution.

### 1.1 Genetic diseases

Using whole genome sequencing, new mutants of genetic diseases were found, which provided new theoretical basis for the pathogenesis of genetic diseases.

(1) Funayama *et al*. (2015) detected by next-generation sequencing technology.A new missense mutation CHCHD2 was found in the genome of 340 independent families with autosomal dominant Parkinson's disease, 517 sporadic Parkinson's disease, and 559 controls. It is speculated that this mutation may be the cause of autosomal dominant Park-


doi: 10.18063/gds.v2i1.

inson's disease.

(2) Nemirovsky *et al*. (2015) genome-wide sequencing of three autism spectrum disorder (ASD) patients and their parents and the use of the most advanced integrated bioinformatics to analyze transmission pathways and prioritize the location of mutations in the gene and the possible relationship with ASD, and Sanger sequencing was used to test them. Confirmation test results showed that exon 21 of SHANK3 gene was missing a heterozygous cytosine chimeric, resulting in a missense sequence of five codons after the termination codon. This mutation may be associated with autism spectrum disorder.

(3) Using the second-generation sequencing technique, the researchers sequenced the entire genome of a soybean caryophyllum ecological model mutated by ethyl methylsulfonate and analyzed its third-generation offspring. A low percentage of prejudiced mutations G/C to A/T (Mohd) were found in each single nucleotide at 208 KB (AS) and 202 KB (AM) intervals. -Yusoff *et al*., 2015). Therefore, genome-wide sequencing of genetic diseases provides a new molecular approach to the etiology of diseases and provides a new direction for the prevention and treatment of these diseases.

## 1.2 Tumor

Tumor is a hotspot and difficult point in current research. Genome-wide sequencing has made great contributions to the discovery, risk factors, prognosis, treatment and drug resistance mechanism of many tumors, and laid a solid foundation for the search for the etiology and radical treatment of tumors.

(1) MRoz *et al*. (2015) calculated the tumor mutation allele heterogeneity of 305 head and neck squamous cell carcinoma patients (clinical and genome-wide sequencing data obtained from the Tumor Genome Project in October 2013) using genome-wide sequencing results. High mutation allele tumor heterogeneity was found to be substantially associated with reduced overall survival Hazard ratio analysis: hazard ratio is high / low heterogeneity =2.2 (95% CI1.4-3.3).

(2) Genome-wide sequencing and exome-wide sequencing of 252 untreated GBM patients from the cancer genome map, 21 GBM patients with first pair recurrence and 2 GBM patients with second pair recurrence revealed that the p53 gene pathway was altered by a high number of subgroups of malignant gliomas. Clonal mutation is a major predictor of molecular events, but can be broadly classified as linear recurrence and sporadic recurrence. Linear recurrence shares extensive genetic similarities with primary tumors and can be directly traced back to one of the characteristics of primary tumors. Branched cells occur, which suggests that genetic changes in primary tumors affect subsequent tumor cell evolution and subclonal heterogeneity (Kim *et al*., 2015).

(3) the researchers analyzed 104 new cases by whole genome sequencing.

Clinical findings from 168 patients with prostate cancer and 168 elderly patients with renewed primary prostate cancer showed that DNA copy num-ber alteration (CAN) load (the percentage of the cancer genome affected by DNA copy num ber changes) was associated with recurrence of primary prostate cancer. The results demonstrated that low-cost use was associated with recurrence of primary prostate cancer. Genome-wide sequencing needle biopsy examines the CAN load to prepare for the prognosis of conservative treatment groups (Hieronymus *et al*., 2014).

(4) Fernandez-Banet et et et *et al*. (2014) genome sequencing of 88 HCC tissues from Chinese patients and adjacent non-tumor liver tissues revealed 4314 rearrangements, including insertion, deletion, inversion and translocation. Two recurrence fusion genes ABCB11 and LRP2 were also found.

(5) Researchers using genome-wide sequencing of serum from patients with pancreatic cancer find that their serum spans genomic DNA across all chromosomes. Serum genome-wide sequencing can be used to identify DNA mutations in patients with pancreatic cancer, and to predict, treat, and assess treatment resistance (Kahlert *et al*., 2014) ).

With the development of various diagnostic techniques, the detection rate of tumors is increasing. However, the specific etiology of tumors has been a worldwide problem. A large number of studies have found that the use of whole genome sequencing has found a lot of tumor pathogenesis, which provides a new theoretical basis for the treatment and prevention of cancer.

## 1.3 Infectious diseases

Infectious diseases are very common, even without enough attention. Infectious diseases are caused by bacteria, viruses, mycoplasma, chlamydia, rickettsia, spirochetes, fungi, parasites and so on. At present, genome-wide sequencing is mainly used to study the mechanism of drug resistance in bacterial infections.

(1) Li *et al*. (2014) Comparing the genome of Staphylococcus capitatum subspecies LNZR-1 with that of five other Staphylococcus capitatum subspecies strains, a large number of biotin resistance genes were found in the subspecies LNZR-1.

(2) Genome-wide sequencing of cyanobacterial ester-resistant Fusarium graminearum strain YP-1, other cyanobacterial ester-sensitive Fusarium graminearum strains and other cyanobacterial ester-resistant Fusarium graminearum strains revealed that the gene encoding myosin-5 (point mutation 216, 217, 4) was present in the resistant strains. 18,420,786) mutation, homologous double-exchange sensitive strain and drug-resistant strain myosin-5 locus, susceptible bacteria and drug-resistant bacteria showed the opposite performance, it can be seen that myosin-5 mutation caused Fusarium graminearum to insecticide cyanide ester resistance (Zheng *et al*., 2015).

(3) Xiao Qiancheng *et al*. (2014) Sequencing of the whole genome of Streptomyces metallococcus revealed that the genes encoding transglutaminase-proactivator-related proteins were identified and the secondary metabolite synthesis gene clusters were predicted, which provided basic data for the functional genomics research of Streptomyces metallococcus.

With the abuse of antibiotics leading to the resistance of many strains to antibiotics, many patients have lost their lives because of drug resistance. However, the specific mechanism of drug resistance is still unclear. The use of genome-wide sequencing technology provides a new detection method for discovering the mechanism of drug resistance, which is a big step forward to solve the problem of drug resistance.

## 1.4 Communicable epidemics

Infectious diseases refer to infectious diseases caused by pathogenic microorganisms that can cause epidemics under certain conditions. Genome-wide sequencing is widely used in epidemiological investigation of malaria, tuberculosis and other infectious diseases, and in the study of drug resistance mechanism.

(1) Logue *et al*. (2015) analyzed the whole genome of the blood or tissues of four species of malaria mosquitoes caught outside Papua New Guinea. Historic mutation studies using homologous DNA sequences larger than 50 Mb to assess contemporary gene flow and population size suggested that despite their morphological, behavioral and ecological similarities It is their group of malaria mosquitoes that evolved independently in the same area.

(2) Whole genome sequencing and partial genome sequencing and genotyping of 236 HBV positive samples from two of Pakistan's largest provinces (Punjab, Sindh) revealed that genotype D subtype D1 was the most prevalent strain of HBV; the most common mutation among HBV carriers was 8 bp deletion mutation (4 out of 23 genotype D samples showed C promoter region deletion in nucleotide 1763-1 774 (n = 2), 1 766-1 773 (n = 1) and 1 767-1 774 (n = 1) 8 bp), A1915T mutation was detected in all samples, and G1679A mutation (86.96%) was detected in genotype D samples (Ahmed *et al*., Ahmed *et al*. 2009).

(3) Walker *et al*. (2014) used genome-wide sequencing in Oxfordshire to investigate the epidemiology of tuberculosis transmission in an unselected population during 2007-2012 and found that people born in low-incidence countries were more likely to develop pulmonary disease (adjusted odds of this association were 1.8, 95% CI 1.2-2.9, P = 0.009), social risk factors 4, 9 5% CI 2.0~9.4, P < 0.0001. Systematic barriers for new entrants may further improve TB control, most importantly for everyone who still needs good health care, especially for high-risk groups.

(4) Genome-wide sequencing of 47 pairs of patients with tuberculosis (one pair was previously treated; the other either failed at 17 weeks of treatment or had a relapsing infection) compared the number and location of single nucleotide polymorphisms in each pair of patients, and found that genome-wide sequencing was useful in differentiating between relapsed and reinfected patients. With greater resolution, it provides additional clear definitions of endpoints for clinical trials (Bryant *et al*., 2013).

(5) Witney *et al*. (2015) In London Teaching Hospital from 2008 to 2014, genome-wide sequencing of 6 suspected highly drug-resistant TB patients revealed that M gene mutations in Mycobacterium tuberculosis were associated with

drug resistance.

Infectious diseases have always been a hidden killer endangering people's health, then, the prevention and control of infectious diseases has become particularly important, the use of genome-wide sequencing for the prevention, diagnosis and treatment of infectious diseases to provide new ideas.

### 1.5 Identify individual disease susceptibility

Whole-genome sequencing of individuals can make patients susceptible to those diseases, which allows individuals to prevent early, reduce the prevalence and improve the quality of life.

(1) Peters *et al*. (2015) found that each embryo caused less than 10 of these errors (a single basic mutation associated with one error per GB has a false positive rate) by genome-wide sequencing of 5-10 cells from two blastocyst-stage embryos, two parents and grandparents. This is the first evidence that genome-wide sequencing can be used. In order to accurately identify new mutations, whole genome sequencing using barcode DNA could be used in the future as part of preimplantation genetic diagnostic methods.

(2) Researchers spent 50 hours in a neonatal intensive care unit in which two babies were clinically differentiated from family genetic disorders by functional automated bioinformatics analysis of the entire genome. They found GJB2-related severe skin disorders in one infant and BRAT1-related fatal newborns in another. Saunders *et al*. (2012).

It can be seen that the discovery of pathogenic gene mutations and the rapid differential diagnosis of diseases by genome-wide sequencing of blastocyst embryos and neonates can reduce the incidence of disease and empirical treatment, and provide faster genetic progress and prognostic counseling.

### 1.6 Biological evolution

Evolution has always been the subject of human exploration, which can find the origin of biological species. Whole genome sequencing can find the theoretical basis of biological evolution by comparing genome structure and using evolutionary analysis, which provides a new method for finding disease-related genes.

(1) Bnyai *et al*. (2014) genome-wide sequencing of a respiratory enterovirus isolated from a green Bush venomous snake revealed that Bush venomous reovirus and hepatoencephalomyelitis virus share several common characteristics, including its genomic structure. Evolutionary analysis showed that the strain, Blum virus and baboon reovirus were monogenic, suggesting that these viruses might have originated from a common ancestor.

(2) Researchers compared the whole genome sequences of Tibetan macaque (M. thibetana, TM), two Chinese rhesus macaques (CR) and two rhesus monkeys (Cynomol gusmacaque, M. fascicularis, CE) and found that TM had 39 million single nucleotide specific mutations. Some genes carry TM-specific homozygous nonsy-mous variants (TSHNVs), and each TSHNVs carries at least 273 immune response and disease-related genes (Fan *et al*., 2014).

(3) Li *et al*. (2014) * using genome sequencing to explore the genetic relationship between European black and white pigs and the races distributed around the globe, we found that many of the non synonymous substitutions single nucleotide polymorphisms loci contained genes into olfactory related categories. Evolutionary analysis revealed a deeper separation between * * pigs between Europe and Asia than between domestic and wild pigs. Analysis of the admixture shows that the native black and white pigs in Europe show a higher portion of the Chinese genetic material, which is consistent with the historical record of its origin. Selective scan analysis showed that strong selection characteristics affected the basic adaptive genomic regions of survival genes: disease resistance, pork production, fertility, meekness and body length. These findings confirm the origin history of European black and white pig * and clarify how to domesticate the pattern of genetic variation through genome wide analysis.

Since Darwin put forward the theory of evolution, human beings have been exploring the mechanism of human evolution, and using whole genome sequencing can provide some new means of exploring the theory of biological evolution.

## 2. Summary and Outlook

In summary, genome-wide sequencing has been widely used in genetic diseases, tumors, infectious diseases, infec-

tious epidemics, judging individual disease susceptibility, biological evolution and so on. It provides a new theoretical basis for disease prevention, pathogenesis, diagnosis and treatment, and provides a new method for human to combat diseases. However, Machini *et al*. (2014) conducted an anonymous survey of 221 genetic counsellors (including genetic counsellors, geneticists, and nurses) about genome-wide sequencing, and found that there were still billing problems, genetic interpretation, and the significance of unknown mutations in genome-wide sequencing. There are still many problems to solve in order to apply genome sequencing to clinical and real life. It is hoped that through human efforts, the whole genome sequencing will eventually be applied to clinical practice, which will facilitate the prevention, diagnosis and treatment of clinical diseases and ultimately improve the quality of life of the people.

# References


1. ahmed c.s., wang z.h., e.g., chen jj, kamal, m., and hou j. hepatitis b virus genotypes, subgenotypes, precore, and basal core promoter mutations in the two largest provinces of pakistan, j. gastroenterol. hepatol., 2009, 24 (4): 569 - 573
2. bányai k., borzák r., ihász k., fehér e., dán 舒., jakab, f., pappt., hetzel u. marschang r.e., and farkas, s.l. whole genome sequencing of a green bush viper reovirus reveal a shared evolutionary history betweenreptilian and unusual mammalian orthoreoviruses, arch. virol 2014, 159 (1): 153 - 163
3. bryant. harris s.r., parkhill j., dawson r., diacon a. h, van helden, p., pym a., mahayiddin a.a., chuchottaworn c., sanne i.m., louw, c., boeree mj, hoelscher m. mchugh t.d., bateson. hunt rd., mwaigwisya p. wright l. gillespie s.h., and bentley s.d. Whole genome sequencing to establish relapse or re infec with mycobacterium tuberculosis and a retrospectiveobser - vational study, lancet respir. med 2013, 1 (10): 786 - 792
4. fernandez banet, j., lee, n.p., chan, k.t., gao h, liu x, sung v.k., tan w., fan s.t., poon r.t., li p., ching k., rejto per annum, mao, m., and kan for decoding complex pat - terns of genomic rearrangement in hepatocellular carcinoma, genomics 2014; 103 (2 - 3): 189 - 203
5. funayama m. ohe k. amo t., furuya, n., yamaguchi j, we see, li y., ogaki k, ando, m., yoshino is tomiyama h., setting-up - ioka k. hasegawa k, we h. satake w., mogushi k. sasaki, r., kokubo y., kuzuhara s. toda. mizuno, y., uchiyama y. ohno, sir k. and hattori n. chchd2 in utosomal dominant late onset parkinson's dis ease: a genome wide linkage andsequencing study, lancet neurol 2015; 14 (3): 274 282
6. "e. schultz, n., gopalan a. carver b.s., m. t. chang, xiao y., heguy a. huberman, k., bernstein, m., software murali m., r. a., vickers, scardino p.t., sander c. v. taylor, reuter, b.s., and sawyer c.l. copy number al teration burden predicts prostate cancer relapse, proc. natl. acad. sci. u.s. 2014, 111 (30): 11139 - 11144
7. kahlert c. melo s.a., protopopov a., tang j., seth s. cook, m., zhang, j., weltz, j., tian l., futreal a., and kalluri r. the identification of double stranded genomic dna spanning all chromosomes with mutated kras and p53 dna in the serum of patients with exosomes pancreatic cancer j. biol. chem 2014, 289 (7): 3869 adopted
8. kim h., zheng s, amini s.s., virk s.m., mikkelsen t. make d. j., grimsby, j., sougnez c. muller, f., hu j. sloan a.e., cohen m.l. van meir, e.g., scarpace l. laird per week, we instein j.n., lander e.s., gabriel s., getz g. meyerson, m., tian l., barnholtz sloan j.s., and verhaak r.g. the sequencing of genomes and multisector exome prima - ry and post treatment of glioblastoma reveal patterns of tumor evolution, genomes res., 2015. 25 (3): 316 - 327
9. li, m., tian s. yeung ck, meng x, tang q, niu l., wang x.,jin l, ma j., long, k., zhou c, cao y. zhu, l., bai, l., tang g, gu y. jiang, a., li x, li r. Whole genome sequencing of berkshire (european native pig) provides insights into its origin and domestication, sci. also, 2014, 4 (4678): 1 - 7
10. li x, lei song, m., y., gong li, l., h. k., liang and jiang x. whole genome sequence and comparative genomic analysis of multidrug resistant staphylococcus capitis subsp. urealyticus strain lnzr - 1, good pathog 2014, 6 (1): 45.
11. logue k. small s.t. chan er, reimer, l., siba p.m., rooms - one per year, and serre, d. to whole genome sequencing reveal absence of recent geneflow and separate demographer ic histories for anopheles punctulatus mosquitoes in papua new guinea, mol. ecol., 2015, 24 (6): 1263 - 1274
12. machini k., douglas j. braxton a., tsipis, j., and kramer k. genetic counselors' views and experiences with the clinical integration of genome sequencing, j. genet. couns 2014, 23 (4): 496 - 505
13. mohd - yusoff new numbering, ruperao p., tomoyoshi n.a., edwards, d., gresshoff p.m., biswas b., and batley, j. scanning the effects of ethyl methanesulfonate on the whole genomes of lotus japonicus using second generation sequencing anal ysis, g3 (bethesda), 2015, 5 (4): 559 - 567
14. mroz *et al*., tward a.m., smidt r.j., ren y., and rocco jw. intra tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the cancer genome atlas, plos med 2015, 12 (2): e1001786
15. nemirovsky s.i., cordoba, m., zaiat jj, air s.p., vega, p. a., gonzalez - morón d., medina, n.m., fabbro m. s., romero, brun b., revale p., ogara m.f., pecci a., marti m, vazquez, m., turjanski a., and kauffman. whole genome sequencing reveal a de novo shank3 mu works in fam iwl autism spectrum disorder, plos one, 2015, 10 (2): e0116358
16. Peters B.A., Kermani B.G., Alferov O., Agarwal, M. R., Mcel Wain - B.A., Gulbahce N. Hayden, D.M., Zhang Tang y.t. tearle, R.Y., R., Crain B., R. A. Munné Prates, Berkeley, S., and R. drmanac, and Phasing, Detection of

Single Base mutations in biopsies Novo from human in vitro fertilized Embryos byadvanced whole - Genome sequencing, Genome Res 2015 25 (3): 427 - 434
17. Saunders C.J., Miller N.A., Soden S.E., Dinwiddie, D.L., Noll A., alnadi N.A., andraws N. Patterson, M.L., krivohlavek L.A. humphray Fellis, J., S., Saffrey P. Kingsbury, Z., J. J. Weir, Betley grocock Margulies, R.J., E.H., e.g. Artman Farrow, M., Safina N.P., petrikin J.E. Hall, K.P., and ki Ngsmore S.F., Rapid Whole Genome sequencing for Genetic disease diagnosis in Neonatal Intensive Care Units, SCI. Transl. Med 2012, 4 (154): 154ra135