



Emerging Era of Exercise Genomics

Seung Kyum Kim PhD

Department of Sports Science, Seoul National University of Science and Technology, Seoul, Korea

Cardiorespiratory endurance has been related to improved cardiovascular health and a longer lifespan. This concept was introduced over 50 years ago and has now been validated by modern science. In 1996, Blair and colleagues reported one of the first large-scale epidemiological data that studied the role of cardiorespiratory fitness as a preventive medical tool after evaluating 10,224 men and 3,120 women for maximal treadmill exercise capacity and following their health for an average of eight years [1]. After controlling for confounding factors such as sex, age, smoking, and blood pressure, the authors discovered that decreased endurance is a major risk factor for all-cause death. This large-scale study is just one of the numerous studies that provide evidence that endurance capacity is a powerful indicator of health and longevity.

As a result of many efforts by exercise physiologists, it is now widely accepted that regularly repeated moderate-intensity exercise, known as aerobic exercise training, can improve cardiorespiratory endurance and thus contribute to a reduction in cardiovascular disease prevalence, morbidity, and mortality. As a result of the numerous positive impacts of exercise on human health, the term “Exercise is Medicine (EIM)” has been coined. EIM originally represented a non-profit initiative launched in 2007 by the American College of Sports Medicine and the American Medical Association and the main purpose of this initiative was to include physical activity as a standard part of medical treatment and patient care process. Nowadays this concept is being expanded and used widely to include preventive methods in medical and exercise science field in order to emphasize the myriad beneficial effects of exercise on human health.

However, it is evident that everyone’s endurance capacity is varied. Furthermore, not everyone responds to exercise in the same way. Previous research has shown that multifactorial effects such as sex, age, an-

thropometric parameters (such as body weight and height), individual lifestyle, baseline health state, and others have a role in exercise capacity and responses to exercise. Another factor that has recently gotten a lot of attention is genetic predisposition.

Accumulating data provide evidence that adherence to a standardized exercise training does not guarantee improvements in fitness for all. In general, exercise capacity can be improved by exercise training in most people; however, there must be individuals showing minimal or no improvements after training and even some people showed negative outcomes. These observations led exercise scientists to raise the question, what genetic factors determine the variation in exercise capacity and responses to exercise training. The research field finding the answer for this question is now called ‘Exercise Genetics’ or ‘Exercise Genomics’.

In fact, the terms, ‘Genetics’ and ‘Genomics’, have slightly different meanings. According to the fact sheet of the National Institute of Health National Human Genome Research Institute (<https://www.genome.gov/about-genomics/fact-sheets/Genetics-vs-Genomics>), ‘Genetics’ refers to the study of how specific traits or conditions are passed down from one generation to another, mainly focusing on genes and their roles. ‘Genomics’ has more broad meanings incorporating the study of a person’s genome and thus primarily incorporates the study of complex diseases with the consideration of genetic and environmental factors. However, they have been used indiscriminately in the exercise science study field in that they have the same purpose of elucidating the genetic factors underlying exercise traits/phenotypes and their responses to exercise training.

To analyze the research trend quantitatively in Exercise Genetics/Genomics, the Web of Science was searched for articles including two words, ‘exercise’ and ‘genetic (OR) genomic’, in the title, abstract, or key-

Corresponding author: Seung Kyum Kim **Tel** +82-2-970-6299 **Fax** +82-2-972-9763 **E-mail** skkim7@seoultech.ac.kr

Keywords Exercise, Single nucleotide polymorphism, Genome-wide association study, Research trend, Co-occurrence

Received 17 Nov 2021 **Revised** 24 Nov 2021 **Accepted** 24 Nov 2021

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

words of publication. It was found that the total number of original articles, excluding reviews, meeting abstracts, editorial, book chapters, and letters, published so far (November 11th, 2021) is 9,384 (Fig. 1). The first manuscript appeared in 1974; however, only a few more articles (n=15) had been published until 1990 (Fig. 2). Whereas, two digits in the number of publications began to be published each year from 1991 and the total number of publications including those two words between 1991 and 2000 was 1,037. This number was a 6,813% increase in comparison to the number of publications (n = 15) between 1974 and 1990, indicating that the new era of Exercise Genetic/Genomics has begun at the beginning of the 1990s. Thereafter, the number of Exercise Genetics/Genomics studies has tended to be more than double each decade so far. The number of original articles published in 2001-2010 and 2011-2021 (present) was 2,752 and 5,595, respectively.

To understand the research trend of Exercise Genetics/Genomics qualitatively, keywords from articles were further analyzed via bibliometric networks using VOSviewer [2]. To classify the research trend by era, keywords were grouped by decades (1991-2000 vs. 2001-2010 vs. 2011-2021) Table 1 exhibits the top 30 keywords most frequently appeared in the articles during three different decade groups. For all decades, general terms commonly used in exercise physiology studies, such as physical activity, risk-factors, skeletal muscle, blood pressure, body mass index, and bone mineral density, were very frequently appeared in the literature. In addition, diseases, such as ‘insulin resistance’, ‘diabetes’, ‘coronary heart disease’, ‘hypertension’, and ‘Alzheimer’s disease’, were the keywords frequently included in the publication. Terms related to Genetics or Genomics, such as ‘genes’, ‘single nucleotide polymorphisms (SNP)’, and ‘mutations’, were also frequently used in the literature for all

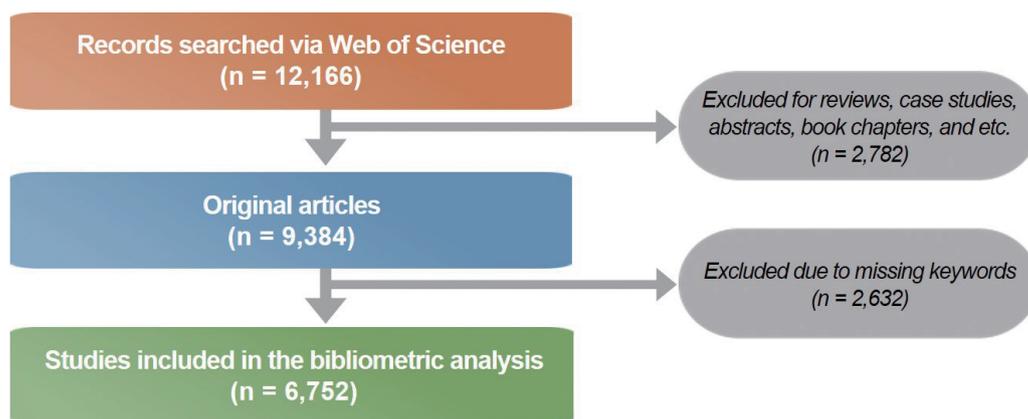


Fig. 1. The procedure of collecting previous studies queried in the Web of Science with two words, ‘exercise’ and ‘genetic OR genomic’, in the title, keywords, or abstract and excluding criteria for the further analyses.

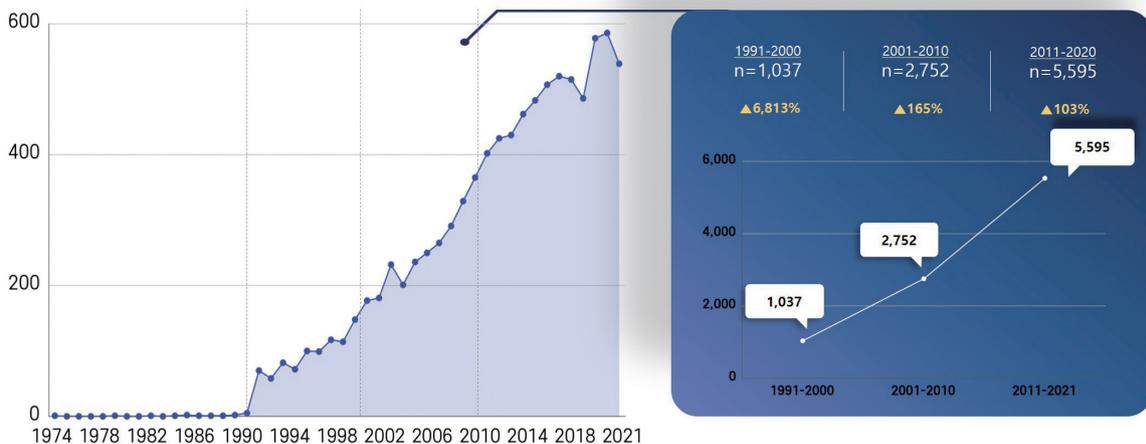


Fig. 2. The number of original articles published each year from 1974 to date and each decade (1991 to 2000 vs. 2001 to 2010 vs. 2011 to 2021).

decades. Noticeably, the keyword, ‘twins’, was highly ranked in the 1990s, while this keyword was not appeared in the 2000s. It might be because the paradigm in the genomic study tended to be shifted around these years. In the 1990s, the majority of genomic studies were conducted on twins, comparing certain traits between monozygotic and dizygotic twins. This traditional comparison was able to provide evidence that the specific trait is genetically influenced, albeit it cannot suggest the underlying mechanisms. From the beginning of the 21st century, genomic studies began in earnest to utilize population-based approaches since a swift advance in sequencing technologies could accelerate genome-wide scanning to identify SNP associated with certain traits/diseases/phenotypes in an unrelated large population. In particular, with

the development of technology that can analyze large-scale SNP on the microarray chip, it became possible to determine the genotype of hundreds of thousands of SNPs at a low cost and in a short time. Genome-Wide Association Study (GWAS) can be now performed to select genetic mutations affecting traits/phenotypes within a large population on statistical significance among hundreds of thousands of SNPs on the chip.

Although little behind, it has been established that this update applied to Exercise Genetics/Genomics as well, as indicated in Table 1. The absolute number of ‘SNP’ appearances is constantly increased over the decades and two new keywords, ‘association’ as a part of GWAS and ‘genome’, were frequently appeared during 2011-2021. In my opinion, the keyword, ‘SNP’ was differently utilized in 1991-2000 compared to later

Table 1. Top 30 keywords appeared in the previous exercise genetics articles each decade

No.	1991-2000			2001-2010			2011-2021		
	Keyword	Freq	Link strength	Keyword	Freq	Link strength	Keyword	Freq	Link strength
1	Physical-activity	118	175	Physical-activity	461	754	Physical-activity	1,138	1,734
2	Risk-factors	80	115	Risk-factors	287	470	Risk-factors	681	1053
3	Genes	70	65	Skeletal-muscle	215	249	Associations	548	962
4	Skeletal-muscle	55	54	Genes	203	242	Skeletal-muscle	431	427
5	Blood-pressure	48	73	SNP	186	295	SNP	424	813
6	Twins	48	78	Genetic	184	276	BMI	338	684
7	BMD	43	73	BMI	158	347	Genes	295	384
8	Diabetes	47	77	Children	156	245	Mutations	288	213
9	Proteins	42	26	Mutations	134	179	Insulin-resistance	205	386
10	Populations	38	49	Insulin-resistance	118	237	Mouse model	202	184
11	SNP	37	38	Mouse	109	117	Diabetes	159	341
12	Mutations	34	33	Blood-pressure	99	168	Life-style	129	277
13	Insulin-resistance	31	55	Diabetes	84	185	Validation	118	159
14	Rat	31	22	BMD	71	116	AD	114	176
15	Sequences	29	19	Aging	68	118	Mitochondria	49	59
16	Asthma	9	5	Life-style	59	131	Survival	49	60
17	BMI	9	21	Rat	55	60	Therapy	49	52
18	Body-composition	9	18	Aerobic capacity	19	23	Genome	48	50
19	Bone mass	9	23	Body composition	19	30	Twins	48	81
20	Calcium intake	9	19	CVD	19	29	Adiposity	47	115
21	CHD	9	14	Cloning	19	16	Anxiety	47	57
22	Density	9	5	Cohort	19	34	Differentiation	47	34
23	Exercise intolerance	9	12	Drosophila	19	12	Genetic	47	79
24	Genetic variation	9	14	Energy-intake	19	48	PMW	47	93
25	Glucose-tolerance	9	17	Hypertension	19	33	Transcription	47	38
26	Hereditiy	9	21	Exercise capacity	19	22	Breast cancer	46	83
27	MRS	9	6	Glucose	19	35	Energy-expenditure	46	78
28	PMW	9	20	Hypoxia	19	20	SCD	46	37
29	Weight-loss	9	22	Loci	19	38	Selection	46	36
30	Aging	8	10	PCR	19	18	Atherosclerosis	45	81

Freq, Frequency; BMD, bone-mineral density; SNP, single nucleotide polymorphisms; BMI, body mass index; CHD, coronary heart disease; MRS, magnetic resonance spectroscopy; PMW, postmenopausal women; CVD, cardiovascular disease; PCR, polymerase chain reaction; AD, Alzheimer's disease; SCD, sudden cardiac death.

years. In the 1990s, most researchers explored the genotype effect of one single gene on exercise-related phenotypes. For example, in 1999, Montgomery and colleagues investigated response to exercise training depending on the genotype of angiotensin-converting enzyme (ACE). In this early Exercise Genetics study, they found subjects carrying I-I genotype had a greater anabolic response than those with D alleles [3]. Thereafter, a few more studies examined the single genotype-dependent difference in exercise traits. However, recently accumulated evidence demonstrate a discrepancy in the role of single gene variants in determining exercise traits, hence the contribution of one SNP is still controversial. Rather, given that current studies show the polygenic (meaning many genes involved in) nature of exercise traits nowadays, the keyword ‘SNP’ during recent decades represents multiple SNPs associated with exercise traits. The research consortium that led the discovery of these multiple

SNPs was the HERITAGE Family study (HEalth, Risk factors, exercise Training And Genetics) [4]. This study was initiated in 1995 to attempt to look at the genetic basis of changes and risk factors for various diseases that occurs as a result of exercise training. As of 2021, more than 200 original articles have been published from this study, providing strong experimental evidence for the genetic regulation of exercise-related traits. It is no exaggeration to say that Exercise Genetics/Genomics is divided into before and after the HERITAGE Family study.

Fig. 3 illustrates the co-occurrence of > top 50 keywords each decade. The color represents the cluster to which keywords appear together. In 1991-2000, 57 keywords appeared more than 7 times were analyzed and as a result, four clusters can be represented (Fig. 3A). Those are the effects of exercise on bone health (red), cardiovascular diseases (blue), skeletal muscles (purple), and RNA-protein (green). From 2001 to 2010, 58

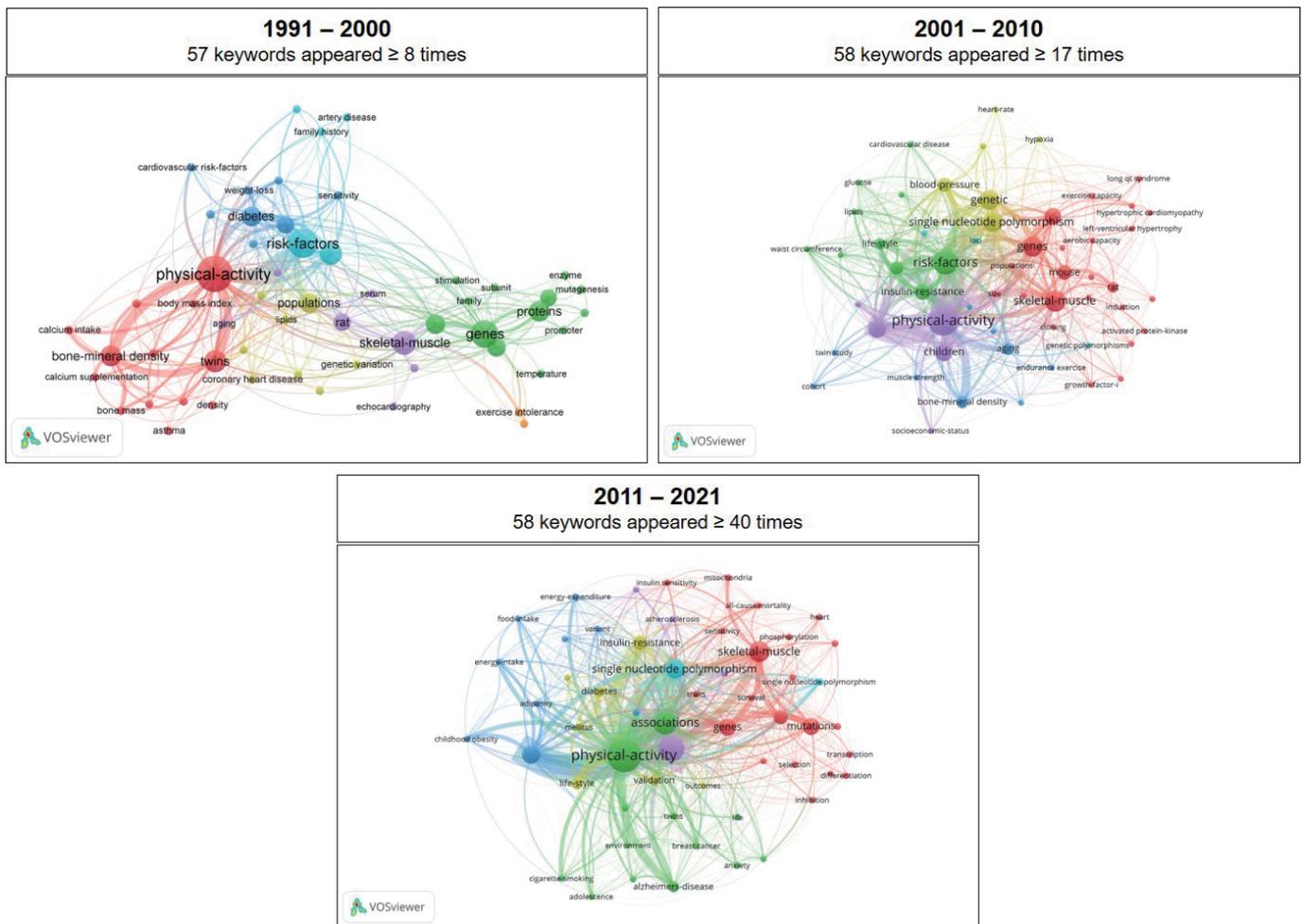


Fig. 3. Co-occurrence analyses for keywords from original articles having two words, ‘exercise’ and ‘genetic OR genomic’, searched in the Web of Science between 1991 to present (2021, November 11th). Both the letter and circle sizes indicate the frequency of appearance, meaning that the higher the frequency of the keyword, the larger the letter and the circle of the keyword. The color is determined by the cluster to which the keyword belongs. Lines represent the link between two keywords, and the thickness of the line demonstrates the frequency of appearance together.

keywords appeared more than 16 times were analyzed (Fig. 3B). For this period, four clusters also can be represented for the effects of exercise on muscle metabolism (red), childhood health (purple) and risk factors for various diseases (green), and genetic components (yellow). For the recent decade (2011-2021), 58 keywords that appeared more than 39 times were analyzed (Fig. 3C). For this decade, three clusters can be distinguished; genetic contribution to the effect of exercise on skeletal muscle (red), the interaction between exercise and environmental factors (green), and the effect of exercise on energy metabolism and obesity (blue). These results demonstrate the different research trends in Exercise Genetics/Genomics each decade. Before 2000, there were relatively many studies investigating the effects of exercise on diseases themselves and the contribution of genes to them. In the 2000s, it showed a propensity of increasing research exploring the genetic influence on the effects of exercise mainly on the risk factors for diseases. For a recent decade (2011–2021), many researchers are highlighting polygenic and multifactorial (i.e. interaction with environmental factors) nature of exercise traits, emphasizing the importance of not only investigating the physiological effects of exercise, but also understanding the genetic basis underlying the exercise physiology to maximize the myriad health benefits we can get from increased physical activity.

Recently, there are emerging terms like ‘trainability’, ‘non-responders to exercise’, and ‘exercise responsive genes’ in the Exercise Genetics/Genomics. ‘Trainability’ refers to the ability of an individual to benefit from exercise training and the variation in trainability across people is being documented. ‘Non-responders to exercise’ describes the persons experiencing no improvements following an exercise training intervention. These suggest that since the persons have low or no trainability possibly due to having low or no ‘exercise responsive genes’, the individualized training regimens rather than the standardized traditional training program should be provided with consideration of their genomic background. Although infrequent, *Exercise Science* journal also has published studies that investigated the genetic contribution to intrinsic exercise phenotypes and/or responses to exercise. Given the fact that exercise plays a significant role as one of the most effective non-pharmaceutical strategies in preventive medicine, efforts to understand the genetic contribution to exercise-related traits (including responses) and unravel the

underlying mechanisms will be high on the future exercise science research agenda. I expect Exercise Science will continue to make a significant contribution to this agenda through diverse research initiatives in the future, as it has in the past. Although the genetic basis for exercise and exercise responses is not fully understood at this time, increasing research efforts should be able to uncover it in the near future. After that, we’ll be able to actually use ‘EIM’ for everyone.

CONFLICT OF INTEREST

The author declares no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, Writing-review & Editing: S.K.Kim.

ORCID

Seung Kyum Kim <https://orcid.org/0000-0002-1708-9393>

REFERENCES

1. Blair SN, Kampert JB, Kohl HW 3rd, Barlow CE, Macera CA, et al. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *JAMA*. 1996;276(3):205-10. PMID: 8667564.
2. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010;84(2):523-538. doi: 10.1007/s11192-009-0146-3. Epub 2009 Dec 31. PMID: 20585380; PMCID: PMC2883932.
3. Montgomery H, Clarkson P, Barnard M, Bell J, Brynes A, et al. Angiotensin-converting-enzyme gene insertion/deletion polymorphism and response to physical training. *Lancet*. 1999;353(9152):541-5. doi: 10.1016/S0140-6736(98)07131-1. PMID: 10028982.
4. Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, et al. The HERITAGE family study. Aims, design, and measurement protocol. *Med Sci Sports Exerc*. 1995;27(5):721-9. PMID: 7674877.