arthritis
research&therapy

REVIEW

Use of next-generation DNA sequencing to analyze genetic variants in rheumatic disease

Graham B Wiley, Jennifer A Kelly and Patrick M Gaffney[*]

Abstract

Next-generation DNA sequencing has revolutionized the field of genetics and genomics, providing researchers with the tools to efficiently identify novel rare and low frequency risk variants, which was not practical with previously available methodologies. These methods allow for the sequence capture of a specific locus or small genetic region all the way up to the entire six billion base pairs of the diploid human genome.

Rheumatic diseases are a huge burden on the US population, affecting more than 46 million Americans. Those afflicted suffer from one or more of the more than 100 diseases characterized by inflammation and loss of function, mainly of the joints, tendons, ligaments, bones, and muscles. While genetics studies of many of these diseases (for example, systemic lupus erythematosus, rheumatoid arthritis, and inflammatory bowel disease) have had major successes in defining their genetic architecture, causal alleles and rare variants have still been elusive. This review describes the current high-throughput DNA sequencing methodologies commercially available and their application to rheumatic diseases in both case–control as well as family-based studies.

Introduction

Within the past 6 years, the advent of high-throughput sequencing methodologies has provided researchers and clinicians with an extremely powerful tool for querying large amounts of the genetic landscape within not only single individuals but also cohorts of many individuals. Often dubbed 'next-generation sequencing' (NGS) or 'second generation sequencing', these methodologies rely on the parallel processing of hundreds of thousands (if not hundreds of millions) of physically sequestered, individually (clonally) amplified copies of DNA, allowing for the generation of massive amounts of data in an extremely short period of time. The resulting datasets, which have become rich gold mines for researchers, provide catalogs of single nucleotide polymorphisms (SNPs), deletion/insertion polymorphisms, copy number variants, and translocations.

NGS DNA methodologies allow researchers to capture particular regions of interest contained within a genome or sequence the entire genome as a whole (whole-genome sequencing). Enriched regions may be specific loci or small genomic regions (targeted sequencing) or the sequences of all known genes and functional elements (exome sequencing). With each method having its own pros and cons, one must consider the scientific objective along with both cost and efficiency when choosing a method. One should not require, for example, the entirety of an exome to be sequenced if the functional variant in question is suspected to be in a non-coding region or previously implicated haplotype block. Similarly, the entire genome need not be sequenced if the study design is focusing only on variants affecting protein-coding genes. Finally, the amount of sequence generated per sample must be taken into account. NGS sequencers are currently optimized to output a set number of reads per run, generally far in excess of a single sample's needs for adequate coverage. To effectively utilize this resource and decrease costs, researchers combine or 'multiplex' samples into shared lanes to reduce cost. This can, however, lead to a decrease in the overall number of reads per sample if the allocation is not meted out judiciously and result in reduced reliability of the calls due to insufficient coverage. Conversely, an overabundance of reads per sample may saturate coverage, diminishing returns on variant calling. Numbers of reads for a given sequence methodology have been empirically ascertained, beyond which increased sequence data yield little or no further variant information [1]. This may increase costs unnecessarily, resulting in fewer samples run for a given budget.

* Correspondence: Patrick-Gaffney@omrf.org
Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK, USA
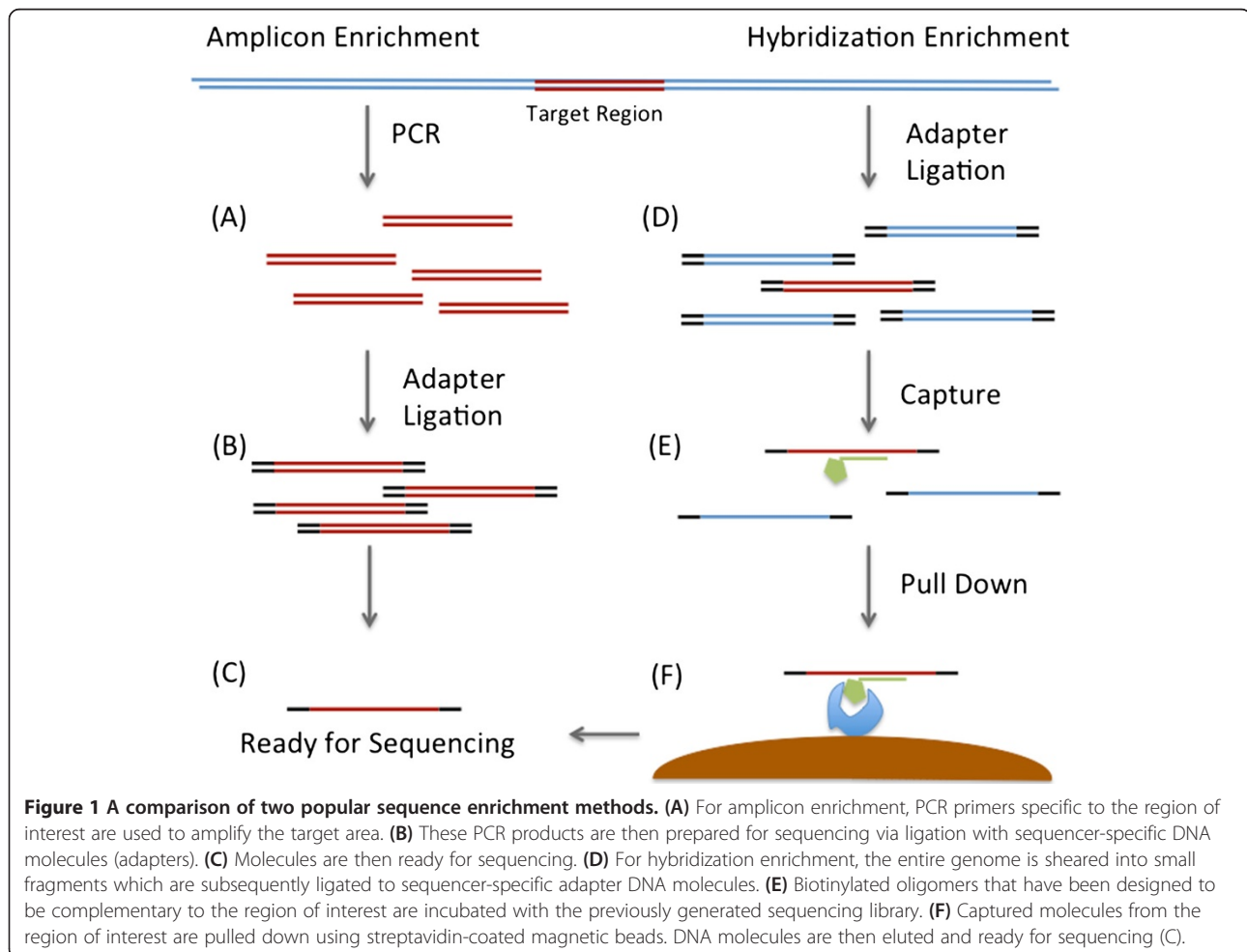
BioMed Central

The major NGS platforms currently available to researchers and clinicians include Illumina's HiSeq and MiSeq, Life Technologies' Ion Torrent and SOLiD, and Roche's 454. While the technologies empowering each of these platforms are quite different, with each having its own nuances in performance and powers of detection, they all rely on the ability to shear DNA into short (<1 kb) fragments, ligate adapters of known sequence to each end, and then immobilize and clonally amplify these molecules onto a solid substrate prior to undergoing massively parallel sequencing. An in-depth discussion of the pros and cons of each technology is beyond the scope of this review, but they are reviewed in other publications [2-4].

Today, these methodologies have revolutionized disease-gene discovery and are now being applied to genetics studies of rheumatic disease. While candidate gene and genome-wide association studies (GWASs) have had great success in identifying candidate genes for many of the rheumatic diseases (for example, >40 known genes in systemic lupus erythematosus (SLE) [5], >100 in rheumatoid arthritis (RA) [6], and >150 in inflammatory bowel disease (IBD) [7]), the extent of heritability explained by the majority of these genes remains small. DNA sequencing methodologies will surely result in additional gene identifications (especially rare variants that are not captured by GWAS methods) that may help explain missing heritability as well as shed light on structural variation within the genome.

High-throughput genomic sequencing methodologies

Targeted sequencing involves the enrichment of a certain locus or group of loci in a varying number of samples. The two most commonly used targeted sequencing approaches are based on either capture with complementary oligomers (hybridization) or amplification via PCR (amplicon) (Figure 1). Hybridization utilizes short biotinylated oligomers that have been designed, generally by an algorithm supplied by the reagent manufacturer, to tile over the locus/loci of interest. These 'bait' oligomers are hybridized to the genomic DNA sample and allow for the capture of their specific complementary DNA sequences. This approach is generally favored for large numbers of loci and has the ability to cover up to 20 million base pairs

Figure 1 A comparison of two popular sequence enrichment methods. (A) For amplicon enrichment, PCR primers specific to the region of interest are used to amplify the target area. **(B)** These PCR products are then prepared for sequencing via ligation with sequencer-specific DNA molecules (adapters). **(C)** Molecules are then ready for sequencing. **(D)** For hybridization enrichment, the entire genome is sheared into small fragments which are subsequently ligated to sequencer-specific adapter DNA molecules. **(E)** Biotinylated oligomers that have been designed to be complementary to the region of interest are incubated with the previously generated sequencing library. **(F)** Captured molecules from the region of interest are pulled down using streptavidin-coated magnetic beads. DNA molecules are then eluted and ready for sequencing (C).

(Mbp) of target regions. Amplicon sequencing methods consist of primer-walking across the locus/loci of interest, followed by pooling the sometimes large number of PCR reactions prior to sequencing. This approach is primarily for regions up to 1 to 2 Mbp total, but allows for large numbers of samples to be pooled together in a single sequencing reaction. Targeted sequencing is often the method of choice for follow-up studies of GWAS associations. Its main disadvantage is that it is generally unable to perform well across repetitive elements within the genome, regions that have low-complexity, or extreme A-T or G-C sequence content.

Exome sequencing is, for all intents and purposes, the same as hybridization targeted capture in methodology. The differences lie in the fact that the exome capture systems have been specifically designed to only capture the coding regions of known genes and, in some cases, known functional non-coding elements of the genome. This optimization allows for a single exome capture system to enrich for 35 to 80 Mbp total. The goal in studying the exomes is to identify mutations that alter the amino acid content of a protein, possibly resulting in altered protein function. Exome capture systems may also include the untranslated regions of genes, pseudogenes, long non-coding RNAs, microRNA genes, and other genomic elements of interest that do not necessarily fall under the moniker of 'gene'. The inclusion of these other loci is heavily dependent on the manufacturer and version of the exome capture system. Since it uses the same methods as targeted sequencing, exome capture technology also shares its disadvantages, with approximately 10% of the exome routinely failing to be captured and, thus, being unable to be sequenced.

Whole-genome sequencing allows for the potential identification of every variant in the genome. It is the most straightforward of the NGS methodologies since the entire genome is prepared and placed onto the sequencer with minimal processing. However, due to the large number of sequencing reads necessary to cover the entire genome, let alone the appropriate amount of coverage necessary to generate good quality variant calls, it remains the most expensive. For this reason very few rheumatic disease studies have yet undertaken whole-genome sequencing. However, we anticipate that this will not be the case for much longer since the cost for whole-genome sequencing continues to decrease.

While we provide below a few examples of how each DNA sequencing methodology has been applied to various rheumatic diseases, additional examples are included for the reader in Table 1.

Other sequencing methodologies

While not a main focus of this review, there are other high-throughput sequencing methods available to researchers that focus on non-genetic variation (epigenetics and transcriptomics). The epigenome consists of alterations resulting from environmental exposures to chemical, nutritional and physical factors that ultimately result in changes to gene expression, suppression, development, or tissue differentiation without altering the underlying DNA sequence. Epigenetic modifications can occur on DNA (methylation) or the histone proteins that compact DNA into nucleosomes (histone modification). Several rheumatic disease studies are already utilizing powerful methods to determine epigenetic influences on phenotype and are discussed in multiple reviews [32-35].

Deep sequencing for transcriptomic studies (RNA-seq) generates more detailed data, including specific isoform, exon-specific transcript and allelic expression levels [36-38], mapping of transcription start sites, identification of sense and antisense transcripts, detection of alternative splicing events, and discovery of unannotated exons [39,40]. To date, RNA-seq methods have been conducted in rheumatic disease studies of RA [41] and SLE [42,43], and in a murine model of inflammatory arthritis [44].

Targeted DNA sequencing approach in rheumatic disease

A number of targeted deep sequencing studies for rheumatic diseases have been used to follow up associations identified by GWASs or custom designed genotyping arrays (Table 1) [25-28]. Adrianto and colleagues [27,28] have performed two such studies in SLE-associated risk loci, *TNFAIP3* and *TNIP1*. *TNFAIP3* was first identified as an SLE risk gene by GWAS and encodes the ubiquitin-modifying enzyme A20, which is a key regulator of NF-kB activity [45,46]. After confirming genetic association in a large case–control association study of five racially diverse populations, Adrianto and colleagues utilized a targeted sequencing approach of the associated *TNFAIP3* risk haplotype in seven carriers (two homozygotes and five heterozygotes) [28]. Though they did not identify any novel SNPs, they did identify a previously unreported single base deletion present on all risk chromosomes. This deletion was adjacent to a rare SNP found in Europeans and Asians and, together, this SNP-indel variant pair formed a TT > A polymorphic dinucleotide that bound to NF-kB subunits with reduced avidity. In addition, the risk haplotype that carried the TT > A variant reduced *TNFAIP3* mRNA and A20 protein expression. *TNIP1* (TNFAIP3 interacting protein 1) has also been associated with SLE in multiple studies, and in conjunction with their studies of *TNFAIP3*, Adrianto and colleagues [27] performed a similar targeted sequencing study of *TNIP1*. Targeted resequencing data resulted in 30 novel variants that were then imputed back into a large, ethnically diverse case–control study, and conditional analysis was used to identify two independent risk haplotypes within *TNIP1* that decrease expression of

Table 1 Rheumatic disease studies utilizing next-generation DNA sequencing methodologies

Disease	Sequencing application	Platform	Study design	Sample size sequenced	Population	Associated genes	Reference
AS	Targeted	Illumina	Case–control	50 cases and 50 controls; 846 cases and 1,308 controls	Han Chinese/ European	*IL23R*	[8]
BD	Exome	Illumina	Case–control	766 cases, 768 controls	Japanese; Turkish	*IL23R, TLR4, MEFV*	[9]
	Exome	Illumina	Case–control	32 cases, 59 controls	Korean	*KIR3DL3, MTHFR, MICA, KIR2KL4, FCGR3A, ICAM1, IFNAR1, KIR2DL4*	[10]
IBD	Exome	Illumina	Case–control	350 cases, 350 controls	Caucasian	*NOD2, IL23R, CARD9, IL18RAP, CUL2, C1orf106, PTPN22, MUC19*	[11]
	Sanger	Applied Biosystems	Case–control	528 cases, 549 controls	Caucasian	*TNFRSF6B*	[12]
CU	Whole genome/Sanger	Illumina/ Applied Biosystems	Family	1 case; 27 cases	Caucasian	*PLCG2*	[13]
DDH	Whole exome	Illumina	Family	4 cases	Caucasian	*CX3CR1*	[14]
FMS	Whole exome	Illumina	Family	19 cases, 150 trios	Caucasian	*ZNF77, KNG1, MMP8, STARD6, C14orf105, FAM81B, C11orf40*	[15]
Gout	Whole genome	Illumina	Case–control	457 cases	Caucasian	*ALDH16A1*	[16]
HUVS	Whole exome/ targeted exome	Illumina	Family	1 case; 3 cases, 9 unaffecteds	Caucasian	*DNASE1L3*	[17]
OA	Sanger	Applied Biosystems	Case–control	992 cases, 944 controls	Caucasian	*GDF5*	[18,19]
	Whole genome	Illumina	Case–control	623 cases, 69,153 controls	Icelandic	*ALDH1A2*	[20]
RA	Targeted	Beckman Coulter	Case–control	34 cases, 74 controls	Alaska Native	*MHC*	[21]
	Exome	Illumina	Case–control	500 cases, 650 controls	Caucasian	*CD2, IL2RA, IL2RB*	[22]
	Whole exome	Illumina	Case–control	19 cases	Japanese	*BTNL2*	[23]
	Whole exome/ targeted exome	Illumina	Consanguineous family/case–control	4 cases; 1,088 cases, 1,088 controls	Caucasian	*PLB1*	[24]
SLE	Targeted	Illumina	Case–control	74 cases, 100 controls	Caucasian	*UBE2L3*	[25]
	Targeted	Illumina	Case–control	100 cases, 100 controls	Caucasian	*IKBKE, IFIH1*	[26]
	Targeted	Illumina	Case–control	218 cases, 157 controls	Caucasian; African-American; Hispanic	*TNIP1*	[27]
	Targeted	Illumina	Case–control	7 cases	Caucasian	*TNFAIP3*	[28]
	Exome	Sanger	Case–control	191 cases, 96 controls	Caucasian	*FAM167A; BLK*	[29]
	Whole exome/Sanger	Life Technologies	Family	1 case; 3 cases, 3 unaffecteds	Caucasian	*PRKCD*	[30]
pSS	Targeted	Sanger	Case–control	19 cases	Caucasian	*TNFAIP3*	[31]

AS, ankylosing spondylitis; BD, Behçet's disease; IBD, inflammatory bowel disease; CU, cold-induced urticaria; DDH, developmental dysplasia of the hip; FMS, fibromyalgia syndrome; HUVS, hypocomplementemic urticarial vasculitis syndrome; OA, osteoarthritis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; pSS, primary Sjögren's syndrome.

TNIP1 mRNA and ABIN1 protein. In a similar fashion, S Wang and colleagues [25] conducted a targeted sequencing study of the SLE-associated *UBE2L3* locus in 74 SLE cases and 100 European controls. They identified five novel variants (three SNPs and two indels) that were not present in NCBI dbSNP build 132, one of which was strongly associated with SLE ($P = 2.56 \times 10^{-6}$). The variants were then imputed back into a large case–control dataset, which ultimately led to the identification of a 67 kb *UBE2L3* risk haplotype in four racial populations that modulates both *UBE2L3* and *UBCH7* expression.

C Wang and colleagues [26] explored the variants within and around *IKBKE* and *IFIH1*, genes also previously identified as associated with SLE. These two genes

were targeted using an amplicon long-range PCR-based strategy of exonic, intronic, and untranslated regions in 100 Swedish SLE cases and 100 Swedish controls. In the course of their sequencing, they identified 91 high-quality SNPs in *IFIH1* and 138 SNPs in *IKBKE*, with 30% of the SNPs identified being novel. Putative functional alleles were then genotyped in a large Swedish cohort, which ultimately yielded two independent association signals within both *IKBKE* (one of which impairs the binding motif of SF1, thus influencing its transcriptional regulatory function) and *IFIH1*.

Davidson and colleagues [8] utilized targeted sequencing of the *IL23R* gene to identify rare polymorphisms associated with ankylosing spondylitis in a Han Chinese population. Targeted sequencing of a 170 kb region containing *IL23R* and its flanking regions was performed in 100 Han Chinese subjects and again in 1,950 subjects of European descent and identified several potentially functional rare variants, including a non-synonymous risk variant (G149R) that proved to be associated with the disease.

Exome studies in rheumatic disease

Many studies have resequenced the exomes of candidate genes to identify variants that are likely to influence protein function and, thus, have biological relevance (Table 1) [9-11,22,29]. For example, Rivas and colleagues [11] utilized targeted exome resequencing to query 56 loci previously associated with IBD. They used an amplicon pooling strategy in 350 IBD cases and 350 controls and identified 429 high confidence variants, 55% of which were not included in dbSNP. Seventy rare and low-frequency protein-altering variants were then genotyped in nine independent case–control datasets comprising 16,054 Crohn's cases, 12,153 ulcerative colitis cases, and 17,575 controls, which identified previously unknown associated IBD risk variants in *NOD2*, *IL18RAP*, *CUL2*, *C1orf106*, *PTPN22*, and *MUC19*. They also identified protective variants within *IL23R* and *CARD9*. Their results were among the first to support the growing hypothesis that common, low-penetrance alleles as well as rare, highly penetrant alleles can exist within the same gene. Other studies have taken a whole exome sequencing approach to target and evaluate all known exonic regions throughout the genome [23].

A primary benefit of these DNA methodologies is the ability to capture rare and low frequency variants that, until now, were unknown. With low frequency variants, however, the power of the widely used indirect linkage disequilibrium-mapping approach is low. Therefore, several studies have performed large-scale targeted exome sequencing studies using genetic burden testing, a method that evaluates the combined effect of an accumulation of rare and low-frequency variants within a particular genomic segment such as a gene or exon. Diogo and colleagues [22] applied this strategy to the exons of 25 RA genes discovered by GWAS while utilizing four burden methods and identified a total of 281 variants (83% with minor allele frequency <1% and 65% previously undescribed), with an accumulation of rare nonsynonymous variants located within the *IL2RA* and *IL2RB* genes that segregated only in the RA cases. Eleven RA case–control dense genotyping array datasets (ImmunoChip and GWAS) comprising 10,609 cases and 35,605 controls were then scrutinized for common SNPs that were in linkage disequilibrium with the 281 variants identified by the exome sequencing. Sixteen of 47 identified variants were subsequently associated with RA, demonstrating that, in addition to previously known common variants, rare and low-frequency variants within the protein-coding sequence of genes discovered by GWASs have small to moderate effect sizes and participate in the genetic contribution to RA. Kirino and colleagues [9] also utilized burden testing while studying the exons of 10 genes identified through GWAS that were associated with Behçet's disease and 11 known innate immunity genes in Japanese and Turkish populations. They used three different burden tests and were able to identify a statistically significant burden of rare, non-synonymous protective variants in *IL23R* (G149R and R381Q) and *TLR4* (D299G and T399I) in both populations, and association of a single risk variant in *MEFV* (M694V) within the Turkish population.

Whole-genome sequencing in rheumatic disease

Until only recently, whole-genome sequencing was an unrealistic option for most studies due to its high costs. Today, however, with a cost approaching $1,000 per sample [47], genetics and genomics researchers are finally able to see this method as a valid option for their studies. To date, few published large-scale whole-genome sequencing studies have been conducted on a rheumatic disease. Sulem and colleagues [16] carried out the first such study, sequencing 457 Icelanders with various neoplastic, cardiovascular and psychiatric conditions to an average depth of at least $10\times$ and identified approximately 16 million variants. These variants were then imputed into a chip-genotyped dataset of 958 gout cases and >40,000 controls with more than 15,000 of these subjects also having measured serum uric acid levels. When analyzing gout as the phenotype, two loci reached genome-wide significance: a novel association with an exonic SNP in *ALDH16A1* ($P = 1.4 \times 10^{-16}$), and a Q141K variant within *ABCG2* ($P = 2.82 \times 10^{-12}$), a gene previously reported to be associated with gout and serum uric acid levels. The *ALDH16A1* SNP displayed stronger association with gout in males and was correlated with a younger age at onset. Four loci reached genome-wide significant association when evaluating association with serum uric acid levels:

the same *ALDH16A1* SNP found with gout ($P = 4.5 \times 10^{-21}$), a novel association with the chromosome 1 centromere ($P = 4.5 \times 10^{-16}$), as well as previously reported signals at *SLE2A9* ($P = 1.0 \times 10^{-80}$) and *ABCG2* ($P = 2.3 \times 10^{-20}$). Another study, by Styrkarsdottir and colleagues [20], utilized whole-genome sequencing of an Icelandic population to further inform a GWAS investigating severe osteoarthritis of the hand. In this case, the imputation of 34.2 million SNPs identified via whole-genome sequencing of 2,230 Icelandic subjects into a previously performed GWAS of 632 cases and 69,153 controls allowed the researchers to identify association with 55 common (41 to 52%) variants within a linkage disequilibrium block containing the gene *ALDH1A2* and four rare (0.02%) variants at 1p31. Other rheumatic disease studies have conducted much smaller scale whole-genome sequencing in one to five individuals followed by targeted exome or Sanger sequencing of the identified variants in larger samples [13].

DNA sequencing in families with rheumatic disease

For rheumatic diseases showing an autosomal dominant or Mendelian inheritance pattern, the study of each genome across multiple generations of the same family can shed light on the variant(s) or gene(s) responsible for disease. Therefore, high-throughput DNA sequencing studies are not limited just to disease cases and population controls, but have been applied to family studies as well [13,14,17,24]. Okada and colleagues [24] recently applied whole-exome sequencing to a four-generation consanguineous Middle Eastern pedigree in which 8 of 49 individuals (16.3%) were affected with RA, which was much higher than the prevalence of RA in the general Middle Eastern population (1%). By applying a novel non-parametric linkage analysis method to GWAS data that looked for regional IBD stretches with a loss of homozygous genotypes in affected cases, they identified a 2.4 Mb region on 2p23 that was enriched in the RA cases. Whole-exome sequencing of 2p23 was performed in four RA cases, which identified a novel single missense mutation within the *PLB1* gene (c.2263G > C; G755R). Variants near the *PBL1* gene were then evaluated in 11 GWAS datasets of 8,875 seropositive RA cases and 29,367 controls, which identified two independent intronic mutations that, when evaluated as a haplotype, demonstrated significant association with RA risk ($P = 3.2 \times 10^{-6}$). Finally, deep exon sequencing of *PBL1* was performed in 1,088 European RA cases and 1,088 European controls, and burden testing revealed an enrichment of rare variants within the protein-coding region of *PBL1.* Taken together, these results suggest both coding and non-coding variants of *PBL1*, a gene that encodes both phospholipase A1 and A2 enzymatic activities, contribute to RA risk.

A major benefit of utilizing NGS methods within families is that researchers are now able to combine previously generated linkage information with new sequence data to identify rare causal variants that contribute to previously detected linkage signals.

Ombrello and colleagues [13] integrated NGS data with previously generated linkage data in three families with a dominantly inherited complex of cold-induced urticaria, antibody deficiency, and autoimmunity. Previous linkage analysis identified a 7.7 Mb interval on chromosome 16q21. Whole-genome sequencing of one affected individual from the first family did not identify any novel mutations within the linkage peak. When analyzing a second family, however, a segregated haplotype containing 24 genes overlapped a linkage interval, and *PLCG2* was subsequently chosen as the most likely candidate. Sequencing of *PLCG2* within family 1 identified a 5.9 kb deletion of exon 19 that was present only in the affected individuals. A *post hoc* analysis of the whole genome data from the family 1 individual confirmed the presence of this deletion. Subsequent sequencing of this gene in the other two families identified further deletions: transcripts in family 2 that lacked exons 20 to 22 because of an 8.2 kb deletion, and deletion of exon 19 in family 3 because of a 4.8 kb deletion. Each of the three deletions affected the carboxy-terminal Src-homology 2 (cSH2) domain of *PLCG2*, a domain that, in healthy individuals, couples the enzymatic activity of *PLCG2* to upstream pathways. In these individuals, however, the deletions resulted in auto-inhibition and constitutive phospholipase activity.

Sanger sequencing in rheumatic disease

Until the application of NGS, Sanger sequencing, which was developed in 1977, was the most widely used sequencing method. However, the advent of NGS does not necessarily ring the death-knell for Sanger sequencing for one or a handful of variants. While on the wane as a large-scale experimental technique, this tried and true methodology still retains usefulness and economy in large-scale replication and screening assays. Many still consider this method to be the 'gold standard' and will utilize Sanger sequencing to validate the results generated by their high-throughput sequencing methods [20,23,24,30]. In addition, recently published studies have applied no other method but Sanger sequencing for deep sequencing of extremely specific regions in smaller numbers of samples. These include a search for rare variants across *GDF5*, a gene harboring a known susceptibility variant for osteoarthritis in 992 cases and 944 controls [18,19], a similar rare variant screen focused on *TNFRSF6B* in pediatric-onset IBD [12], exome sequencing of *TNFAIP3* in 19 primary Sjögren's syndrome patients with lymphoma [31], and targeted sequencing of the

FAM167 and *BLK* exomes in 191 SLE cases and 96 controls [29].

The future of sequencing

While a tried and true advancement in the genetics and genomics of rheumatic disease studies, deep sequencing, as a technological field, has and will continue to remain in a state of flux. With the continued refinement of technology and methods, sequencing costs have dropped tremendously over the past 5 years and, as of the drafting of this manuscript, whole-genome sequencing of humans has dropped to less than $1,000 per sample [48]. At this price point, the continued viability of exome sequencing as a widespread technique has yet to be determined. Indeed, it is quite within the realm of possibility that all patients will have their genomes sequenced as a routine test at presentation to their healthcare provider. The foreseeable rise of nanopore sequencers and other 'third-generation' sequencers able to process single molecules of DNA may make bedside sequencing a reality.

Note: This article is part of the series 'New technologies'. Other articles in this series can be found at http://arthritis-research.com/series/technology.

Abbreviations

GWAS: Genome-wide association study; IBD: Inflammatory bowel disease; Mbp: Million base pairs; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; RA: Rheumatoid arthritis; SLE: Systemic lupus erythematosus; SNP: Single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

References

1. Ajay SS, Parker SCJ, Abaan HO, Fajardo KVF, Margulies EH: Accurate and comprehensive sequencing of personal genomes. *Genome Res* 2011, **21**:1498–1505.
2. Jessri M, Farah CS: Next generation sequencing and its application in deciphering head and neck cancer. *Oral Oncol* 2014, **50**:247–253.
3. Rieber N, Zapatka M, Lasitschka B, Jones D, Northcott P, Hutter B, Jäger N, Kool M, Taylor M, Lichter P, Pfister S, Wolf S, Brors B, Eils R: Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS One* 2013, **8**:e66621.
4. Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, Ji HP, Snyder M: Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol* 2012, **30**:78–82.
5. Cui Y, Sheng Y, Zhang X: Genetic susceptibility to SLE: recent progress from GWAS. *J Autoimmun* 2013, **41**:25–33.
6. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su D-F, Yang J, Xie G, Keystone E, Westra H-J, Esko T, Metspalu A, et al: Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014, **506**:376–381.
7. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, Essers J, Mitrovic M, Ning K, Cleynen I, Theatre E, Spain SL, Raychaudhuri S, Goyette P, Wei Z, Abraham C, Achkar J-P, Ahmad T, Amininejad L, Ananthakrishnan AN, Andersen V, Andrews JM, Baidoo L, Balschun T, Bampton PA, Bitton A, et al: Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012, **491**:119–124.
8. Davidson SI, Jiang L, Cortes A, Wu X, Glazov EA, Donskoi M, Zheng Y, Danoy PA, Liu Y, Thomas GP, Brown MA, Xu H: Brief report: high-throughput sequencing of IL23R reveals a low-frequency, nonsynonymous single-nucleotide polymorphism that is associated with ankylosing spondylitis in a Han Chinese population. *Arthritis Rheum* 2013, **65**:1747–1752.
9. Kirino Y, Zhou Q, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Özyazgan Y, Ugurlu S, Erer B, Abaci N, Ustek D, Meguro A, Ueda A, Takeno M, Inoko H, Ombrello MJ, Satorius CL, Maskeri B, Mullikin JC, Sun H-W, Gutierrez-Cruz G, Kim Y, Wilson AF, Kastner DL, Gül A, Remmers EF: Targeted resequencing implicates the familial Mediterranean fever gene MEFV and the toll-like receptor 4 gene TLR4 in Behçet disease. *Proc Natl Acad Sci U S A* 2013, **110**:8134–8139.
10. Kim SJ, Lee S, Park C, Seo J-S, Kim J-I, Yu HG: Targeted resequencing of candidate genes reveals novel variants associated with severe Behçet's uveitis. *Exp Mol Med* 2013, **45**:e49.
11. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N: Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011, **43**:1066–1073.
12. Cardinale CJ, Wei Z, Panossian S, Wang F, Kim CE, Mentch FD, Chiavacci RM, Kachelries KE, Pandey R, Grant S: Targeted resequencing identifies defective variants of decoy receptor 3 in pediatric-onset inflammatory bowel disease. *Genes Immun* 2013, **14**:447–452.
13. Ombrello MJ, Remmers EF, Sun G, Freeman AF, Datta S, Torabi-Parizi P, Subramanian N, Bunney TD, Baxendale RW, Martins MS, Romberg N, Komarow H, Aksentijevich I, Kim HS, Ho J, Cruse G, Jung M-Y, Gilfillan AM, Metcalfe DD, Nelson C, O'Brien M, Wisch L, Stone K, Douek DC, Gandhi C, Wanderer AA, Lee H, Nelson SF, Shianna KV, Cirulli ET, et al: Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. *N Engl J Med* 2012, **366**:330–338.
14. Feldman GJ, Parvizi J, Levenstien M, Scott K, Erickson JA, Fortina P, Devoto M, Peters CL: Developmental dysplasia of the hip: linkage mapping and whole exome sequencing identify a shared variant in CX3CR1 in all affected members of a large multigeneration family. *J Bone Miner Res* 2013, **28**:2540–2549.
15. Feng J, Zhang Z, Wu X, Mao A, Chang F, Deng X, Gao H, Ouyang C, Dery KJ, Le K, Longmate J, Marek C, St Amand RP, Krontiris TG, Shively JE: Discovery of potential new gene variants and inflammatory cytokine associations with fibromyalgia syndrome by whole exome sequencing. *PLoS One* 2013, **8**:e65033.
16. Sulem P, Gudbjartsson DF, Walters GB, Helgadottir HT, Helgason A, Gudjonsson SA, Zanon C, Besenbacher S, Bjornsdottir G, Magnusson OT, Magnusson G, Hjartarson E, Saemundsdottir J, Gylfason A, Jonasdottir A, Holm H, Karason A, Rafnar T, Stefansson H, Andreassen OA, Pedersen JH, Pack AI, de Visser MCH, Kiemeney LA, Geirsson AJ, Eyjolfsson GI, Olafsson I, Kong A, Masson G, Jonsson H, et al: Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat Genet* 2011, **43**:1127–1130.
17. Ozçakar ZB, Foster J, Diaz-Horta O, Kasapcopur O, Fan Y-S, Yalçınkaya F, Tekin M: DNASE1L3 mutations in hypocomplementemic urticarial vasculitis syndrome. *Arthritis Rheum* 2013, **65**:2183–2189.
18. Dodd AW, Syddall CM, Loughlin J: A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *Eur J Hum Genet* 2013, **21**:517–521.
19. Dodd AW, Rodriguez-Fontenla C, Calaza M, Carr A, Gomez-Reino JJ, Tsezou A, Reynard LN, Gonzalez A, Loughlin J: Deep sequencing of GDF5 reveals the absence of rare variants at this important osteoarthritis susceptibility locus. *Osteoarthritis Cartilage* 2011, **19**:430–434.
20. Styrkarsdottir U, Thorleifsson G, Helgadottir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, Strijbosch AM, Evangelou E, Hart D, Beekman M, Jonasdottir A, Sigurdsson A, Eiriksson FF, Thorsteinsdottir M, Frigge ML, Kong A, Gudjonsson SA, Magnusson OT, Masson G, Hofman A, Arden NK, Ingvarsson T, Lohmander S, Kloppenburg M, Rivadeneira F, Nelissen RGHH,

Spector T, Uitterlinden A, Slagboom PE, Thorsteinsdottir U, *et al*: **Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31.** *Nat Genet* 2014, **46**:498–502.

21. Yan Z, Ferucci ED, Geraghty DE, Yang Y, Lanier AP, Smith WP, Zhao LP, Hansen JA, Nelson JL: **Resequencing of the human major histocompatibility complex in patients with rheumatoid arthritis and healthy controls in Alaska Natives of Southeast Alaska.** *Tissue Antigens* 2007, **70**:487–494.

22. Diogo D, Kurreeman F, Stahl EA, Liao KP, Gupta N, Greenberg JD, Rivas MA, Hickey B, Flannick J, Thomson B: **Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis.** *Am J Hum Genet* 2013, **92**:15–27.

23. Mitsunaga S, Hosomichi K, Okudaira Y, Nakaoka H, Kunii N, Suzuki Y, Kuwana M, Sato S, Kaneko Y, Homma Y, Kashiwase K, Azuma F, Kulski JK, Inoue I, Inoko H: **Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the BTNL2.** *J Hum Genet* 2013, **58**:210–215.

24. Okada Y, Diogo D, Greenberg JD, Mouassess F, Achkar WAL, Fulton RS, Denny JC, Gupta N, Mirel D, Gabriel S, Li G, Kremer JM, Pappas DA, Carroll RJ, Eyler AE, Trynka G, Stahl EA, Cui J, Saxena R, Coenen MJH, Guchelaar H-J, Huizinga TWJ, Dieudé P, Mariette X, Barton A, Canhão H, Fonseca JE, de Vries N, Tak PP, Moreland LW, *et al*: **Integration of sequence data from a consanguineous family with genetic data from an outbred population identifies PLB1 as a candidate rheumatoid arthritis risk gene.** *PLoS One* 2014, **9**:e87645.

25. Wang S, Adrianto I, Wiley GB, Lessard CJ, Kelly JA, Adler AJ, Glenn SB, Williams AH, Ziegler JT, Comeau ME, Marion MC, Wakeland BE, Liang C, Kaufman KM, Guthridge JM, Alarcon-Riquelme ME, Biolupus, Networks G, Alarcon GS, Anaya JM, Bae SC, Kim JH, Joo YB, Boackle SA, Brown EE, Petri MA, Ramsey-Goldman R, Reveille JD, Vila LM, Criswell LA, *et al*: **A functional haplotype of UBE2L3 confers risk for systemic lupus erythematosus.** *Genes Immun* 2012, **13**:380–387.

26. Wang C, Ahlford A, Laxman N, Nordmark G, Eloranta ML, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jonsen A, Bengtsson AA, Truedsson L, Rantapaa-Dahlqvist S, Sjöwall C, Sandling JK, Ronnblom L, Syvanen AC: **Contribution of IKBKE and IFIH1 gene variants to SLE susceptibility.** *Genes Immun* 2013, **14**:217–222.

27. Adrianto I, Wang S, Wiley GB, Lessard CJ, Kelly JA, Adler AJ, Glenn SB, Williams AH, Ziegler JT, Comeau ME, Marion MC, Wakeland BE, Liang C, Kaufman KM, Guthridge JM, Alarcon-Riquelme ME, Alarcon GS, Anaya JM, Bae SC, Kim JH, Joo YB, Boackle SA, Brown EE, Petri MA, Ramsey-Goldman R, Reveille JD, Vila LM, Criswell LA, Edberg JC, Freedman BI, *et al*: **Association of two independent functional risk haplotypes in TNIP1 with systemic lupus erythematosus.** *Arthritis Rheum* 2012, **64**:3695–3705.

28. Adrianto I, Wen F, Templeton A, Wiley G, King JB, Lessard CJ, Bates JS, Hu Y, Kelly JA, Kaufman KM, Guthridge JM, Alarcón-Riquelme ME, Anaya J-M, Bae S-C, Bang S-Y, Boackle SA, Brown EE, Petri MA, Gallant C, Ramsey-Goldman R, Reveille JD, Vilá LM, Criswell LA, Edberg JC, Freedman BI, Gregersen PK, Gilkeson GS, Jacob CO, James JA, BIOLUPUS and GENLES Networks, *et al*: **Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus.** *Nat Genet* 2011, **43**:253–258.

29. Guthridge JM, Lu R, Sun H, Sun C, Wiley GB, Domínguez N, Macwana SR, Lessard CJ, Kim-Howard X, Cobb BL, Kaufman KM, Kelly JA, Langefeld CD, Adler AJ, Harley ITW, Merrill JT, Gilkeson GS, Kamen DL, Niewold TB, Brown EE, Edberg JC, Petri MA, Ramsey-Goldman R, Reveille JD, Vilá LM, Kimberly RP, Freedman BI, Stevens AM, Boackle SA, Criswell LA, *et al*: **Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription.** *Am J Hum Genet* 2014, **94**:586–598.

30. Belot A, Kasher PR, Trotter EW, Foray A-P, Debaud A-L, Rice GI, Szynkiewicz M, Zabot M-T, Rouvet I, Bhaskar SS, Daly SB, Dickerson JE, Mayer J, O'Sullivan J, Juillard L, Urquhart JE, Fawdar S, Marusiak AA, Stephenson N, Waszkowycz B, W Beresford M, Biesecker LG, C M Black G, René C, Eliaou JF, Fabien N, Ranchin B, Cochat P, Gaffney PM, Rozenberg F, *et al*: **Protein kinase cδ deficiency causes mendelian systemic lupus erythematosus with B cell-defective apoptosis and hyperproliferation.** *Arthritis Rheum* 2013, **65**:2161–2171.

31. Nocturne G, Boudaoud S, Miceli-Richard C, Viengchareun S, Lazure T, Nititham J, Taylor KE, Ma A, Busato F, Melki J, Lessard CJ, Sivils KL, Dubost JJ, Hachulla E, Gottenberg JE, Lombes M, Tost J, Criswell LA, Mariette X:

32. **Germline and somatic genetic variations of TNFAIP3 in lymphoma complicating primary Sjogren's syndrome.** *Blood* 2013, **122**:4068–4076.

32. Glant TT, Mikecz K, Rauch TA: **Epigenetics in the pathogenesis of rheumatoid arthritis.** *BMC Med* 2014, **12**:35.

33. Gray SG: **Epigenetic-based immune intervention for rheumatic diseases.** *Epigenomics* 2014, **6**:253–271.

34. Zan H: **Epigenetics in lupus.** *Autoimmunity* 2014, **47**:213–214.

35. Costa-Reis P, Sullivan KE: **Genetics and epigenetics of systemic lupus erythematosus.** *Curr Rheumatol Rep* 2013, **15**:369.

36. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57–63.

37. Richard H, Schulz MH, Sultan M, Nurnberger A, Schrinner S, Balzereit D, Dagand E, Rasche A, Lehrach H, Vingron M, Haas SA, Yaspo ML: **Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments.** *Nucleic Acids Res* 2010, **38**:e112.

38. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science* 2008, **321**:956–960.

39. Ozsolak F, Milos PM: **RNA sequencing: advances, challenges and opportunities.** *Nat Rev Genet* 2011, **12**:87–98.

40. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK: **Understanding mechanisms underlying human gene expression variation with RNA sequencing.** *Nature* 2010, **464**:768–772.

41. Heruth DP, Gibson M, Grigoryev DN, Zhang LQ, Ye SQ: **RNA-seq analysis of synovial fibroblasts brings new insights into rheumatoid arthritis.** *Cell Biosci* 2012, **2**:43.

42. Stone RC, Du P, Feng D, Dhawan K, Rönnblom L, Eloranta M-L, Donnelly R, Barnes BJ: **RNA-Seq for enrichment and analysis of IRF5 transcript expression in SLE.** *PLoS One* 2013, **8**:e54487.

43. Shi L, Zhang Z, Yu AM, Wang W, Wei Z, Akhter E, Maurer K, Costa-Reis P, Song L, Petri M, Sullivan KE: **The SLE transcriptome exhibits evidence of chronic endotoxin exposure and has widespread dysregulation of non-coding and coding RNAs.** *PLoS One* 2014, **9**:e93846.

44. Zhang H, Hilton MJ, Anolik JH, Welle SL, Zhao C, Yao Z, Li X, Wang Z, Boyce BF, Xing L: **NOTCH inhibits osteoblast formation in inflammatory arthritis via noncanonical NF-κB.** *J Clin Invest* 2014, **124**:3200–3214.

45. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham DC, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM: **Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40**:1059–1061.

46. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, Shifrin N, Petri MA, Kamboh MI, Manzi S, Seldin MF, Gregersen PK, Behrens TW, Ma A, Kwok P-Y, Criswell LA: **Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40**:1062–1064.

47. **DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).** http://www.genome.gov/sequencingcosts/

48. Hayden EC: **Technology: the $1,000 genome.** *Nature* 2014, **507**:294–295.