Future of genomics in diagnosis of human arthritis: the hype, hope and metamorphosis for tomorrow

Ashok R Amin†, Seth D Thompson & Shailey A Amin

†Author for correspondence Carilion Clinic, 101 Elm Avenue, Research Department, Roanoke, VA 24013, USA Tel.: +1 540 985 8499; Fax: +1 540 344 7278; aramin@carilion.com

The post-genomic era, which is fueled by automation and other technologies, provokes a change in our grossly naive view of genetic determinism (that single genes govern complex traits) to the obvious reality that most human diseases are complex entities. Gene(s), although necessary, contribute only partially to disease, while environmental factors, lifestyles, epigenetics and epistasis significantly influence pathophysiology and, eventually, the expression of transient biomarkers that can be utilized for diagnosis and prognosis. Human osteoarthritis and rheumatoid arthritis are multifactorial, complex diseases. The genetic inheritance of these diseases remains elusive, although they tend to run in families wherein some siblings have a two- to tenfold increased risk of developing the diseases.

Keywords: biomarkers, candidate genes, collagen, complex disease, epigenetic, gene chips, genomics, osteoarthritis, predictive medicine, rheumatoid arthritis

future medicine part of fsg

Hunt for the osteoarthritis/rheumatoid arthritis gene

Monogenic diseases, in contrast to complex diseases, comprise less than 3% of the known human diseases. The candidate-gene approach for osteoarthritis (OA) and rheumatoid arthritis (RA) has yielded gene identities with incoherent follow-up results. The selection of candidate genes has many parallels to the processes of identification and ranking of risk factors in an epidemiological study and selection from mRNA expression arrays of affected tissues. The effort to identify a single dysfunctional OA or RA gene has been based on an *a priori* hypothesis. For example, skeletal dysplasia is a group of heritable diseases that represents monogenic genetic models common in bone and joint diseases, including chondrodysplasia in children. Several skeletal dysplasias are complicated by OA, and these have been viewed as early-onset forms of OA with a genetic component in a single gene or family of genes responsible for expression of matrix materials (such as collagen) involved in the homeostasis of bone and cartilage and/or inflammatory mediators. These genes include *COL2A1*, *COL9A1-A3*, *COL11A* and *-A2*, *SED2*, *MATN3*, *COMP*, *WISP3*, *IL-1* and *TNF* [1]. Such studies typically compare single nucleotide polymorphisms (SNPs) in a lone candidate gene (e.g., the collagen gene; *COL9A3*) or a specific chromosomal region (e.g., the long arm of chromosome 13) in healthy and sick individuals [2]. Phenotypic disparities can arise from the same gene, as exhibited by the *IL-1* cluster that encodes for susceptibility to knee OA but not hip OA, accentuating the need for narrowing the selection of patient phenotypes for genetic/genomic studies. The candidate genes described above were normally inherited as fully penetrant autosomal dominant disorders and conferred OA characteristics early in life rather than in the aging population. Although these findings were encouraging, the associated investigations suffered from small sample size and relatively poor phenotypic characterization (less than 2% of the general OA patient population) and borderline positives. Recent microsatellite studies have failed to support the role of various gene clusters previously linked to RA, such as the provocative inflammatory cytokine cluster on *5q31.1*, *IL-10* and *IRF* genes [3]. The correlation of RA with HLA-DR remains one of the most reproducible findings that shed additional light on the pathophysiology of the disease [4]. The tools and resources used to generate high quality data continue to progress and will aid in focusing sharp images in a blurry genome background.

Scanning the genome & connecting the dots

Genome-wide linkage scans, which make no prior predictions of gene(s) coding for OA or RA traits, and are based on the principal that an individual has the tendency to inherit two or more nonallelic genes by independent assortment of linked markers located proximally on the same chromosome, are an alternative to the candidate gene approach. Two opposing patterns have emerged employing this approach. These are that OA or RA segregate as a Mendelian trait or that a non-Mendelian process is involved. OA exhibits a clear predilection for specific joints: it appears most commonly in the hip, knee joint, lumbar and cervical spine, as well as the distal interphalangeal (DIP), the first carpometacarpal (CMCI; at the base of the thumb) and proximal interphalangeal joints of the hand. Several studies

have suggested that OA of the hand has a significant genetic component where Heberden nodes (osteophytes of DIP joints) and CMCI-joint traits are inherited three- and seven-times more commonly, respectively, in sisters of affected individuals than in the general population. Similarly, the genetic contribution of inheritance has been estimated to be as much as 65% in patients with OA of the hand or knee. The genome-wide approaches have yielded low logarithm of the odds scores, resulting in questionable reliability. Female hip OA showed suggestive linkages to chromosome 2q and 4q; early-onset OA to 6q and 11q; and the common form of OA to 16p and 16q. Linkage analysis of affected sibling pair cohorts and of rare families in which primary OA segregates as a Mendelian trait, has identified regions of suggestive linkages on chromosomes 2q, 4q, 6, 7p, 11q, 12q, 16 and Xcen. One of the new approaches to genome scanning involves characterization of the potential function(s) of 1% of the human genome using a system approach. Briefly, 500,000 genetic markers were analyzed in 2000 patients with various complex diseases, including RA, of which three of the genetic markers linked strongly to RA [5]. These studies will facilitate connecting the dots among multiple genes and across chromosomes to generate a picture for predictive diagnosis/risk.

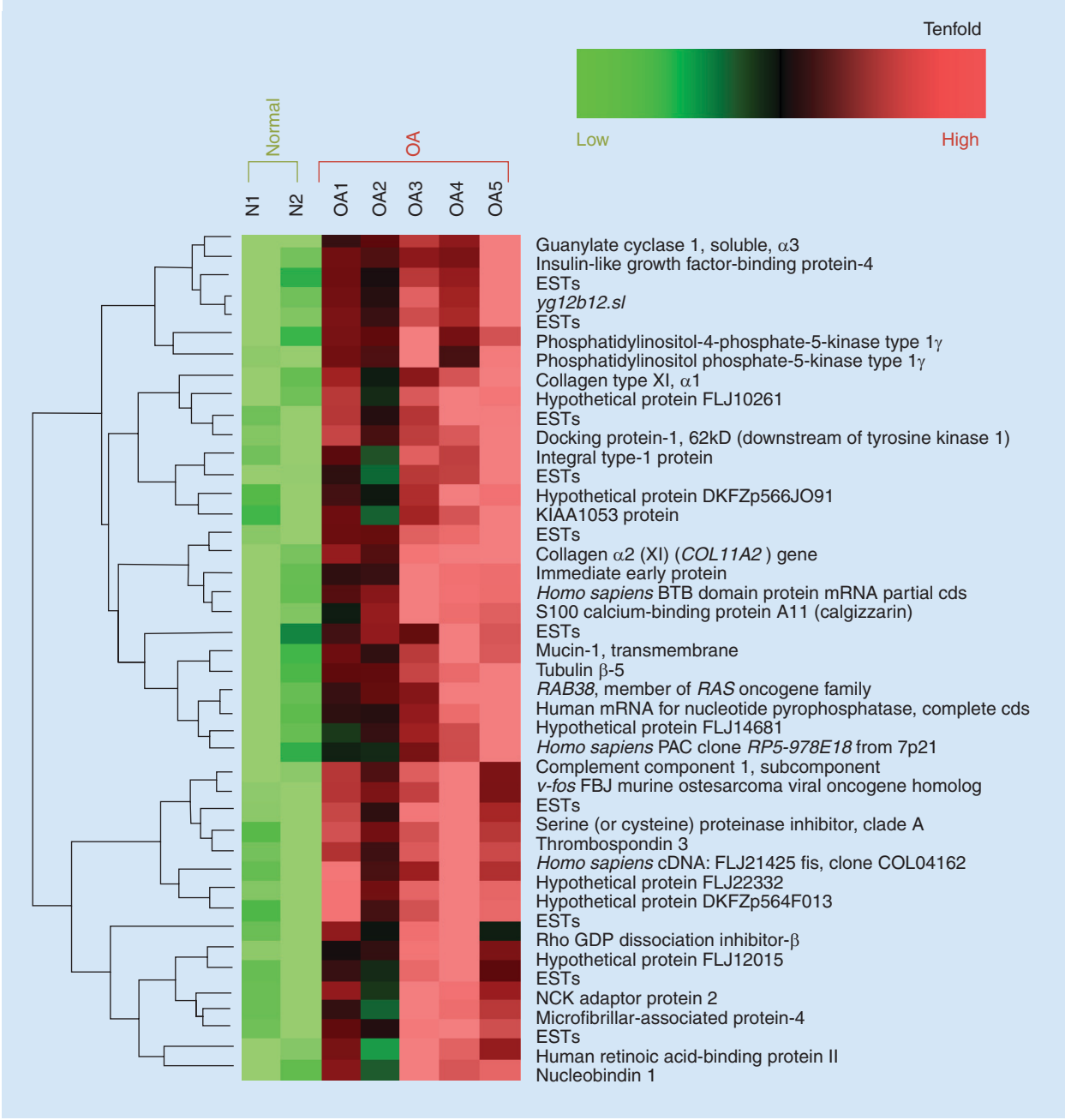
Epigenetic inheritance, silenced genes: a second voice in arthritis?

This emerging concept introduces a monkey wrench in an existing mayhem, but commands consideration. Human monozygotic twins (genetically identical) are often discordant for important phenotypes, including complex diseases. Such variations within virtually identical DNA sequences are largely attributed to the effects of the environment and aging, the latter contributing to an amazing array of aberrations in proliferate homeostasis such as atrophies and hyperplasias, which occur simultaneously within the same tissue (cartilage and synovium). Stochastic epigenetic shifts in gene expression have been observed in both OA and RA [6]. Aberrant epigenetic changes in histone deacetylase expression have been reported in both OA and RA [6,7]. Although the role of epigenetics in arthritis remains elusive, it is tempting to speculate that altered gene expression during aging may be a selectionist adaptive strategy to cope with diverse stochastic environmental and lifestyle changes associated with complex diseases. The jury is still out on this theory.

A unified approach with multiple technologies

Phenotype complexity is one of the major impediments in OA/RA genetics and genomics, as it remains the starting point of these studies [8]. In fact, there is every reason to believe that a condition traditionally labeled as OA or RA will not directly be attributable to a specific gene(s). Thus, OA gene(s) may be associated with some early nonsymptomatic aspects of OA (such as obesity and/or sports injury and/or lifestyle) followed by ovate inflammation (i.e., molecular inflammation without clinical symptoms), a gradual progression towards a symptomatic feature (joint and synovial inflammation), late chronic conditions (cartilage and joint destruction) and a variety of molecular features (upregulation of proteases). A system biology approach, which converges various disciplines of genomics, bioinformatics and proteomics [9] with well curated, phenotypically characterized, clinical samples, will be essential to generate more realistic patterns of the players in human OA. For example, more than 1500 mRNA transcripts that are abnormally expressed in most human OA cartilage as compared with normal control cartilage have been identified by gene arrays (**Figure 1**) and a majority of them have been validated by real time PCR with a separate set of phenotypically similar clinical samples [10]. Moreover, small effects in quantitative traits (such as specific low-abundance mRNA for TNF-α convertase in OA) may escape identification by one approach [11]. Common mRNA profiles have been identified in the cartilage, synovial tissue and peripheral blood cells of OA patients. Gene expression in early RA-peripheral blood mononuclear cells (PBMCs) indicated response to an unknown viral antigen [12]. These observations suggest that, besides the dramatic molecular events in OA/RA-affected joints, there are disease-specific signatures that are systemic in nature; blood PBMCs, which may serve as biomarkers for diagnosis and prognosis. Although the gene array signatures possess complexities depending on gene chip, method of preparation of probe sets or sample, or the level of quality of the clinical sample, this technology demonstrates a singular potential to cover the diversities of OA or RA in a single snapshot. Gene arrays have not only allowed differentiation of OA from RA, but also discriminate subsets of RA with respect to interferon expression [13] or OA and RA (flare and remission) (**Figure 2**).

Figure 1. Gene-expression profile in normal and osteoarthritis-affected cartilage.

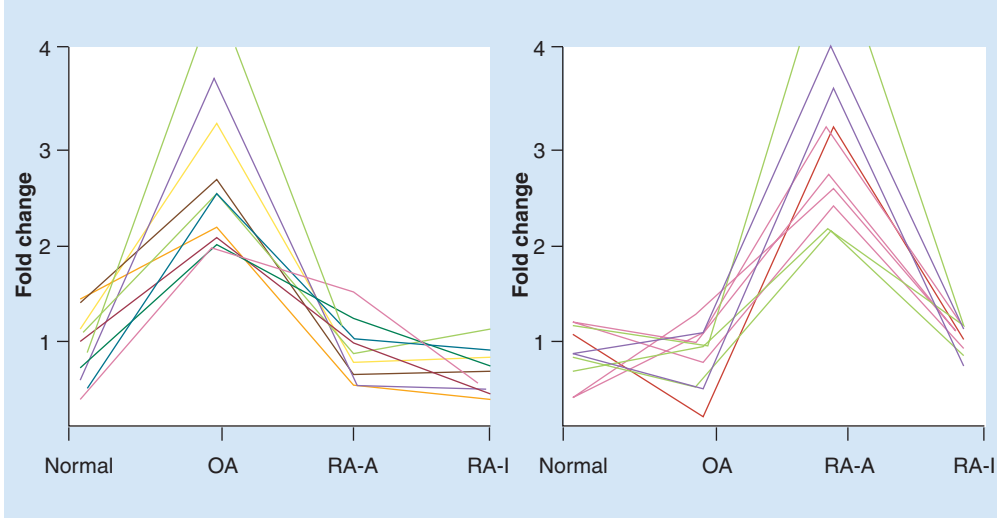


Knee articular cartilages were obtained from normal and OA-affected patients undergoing knee replacement surgery. Total RNA was isolated from cartilage of nonarthritic (n = 20) individuals and pooled into two groups, N1 and N2. Similarly, total RNA was also isolated from OA patients (n = 70), which was pooled (10–15 samples each) into five groups (OA1–OA5). RNA was hybridized against Affymetrix GeneChip®, as described in Attur *et al.* 2002 [9]. Per-chip and per-gene normalizations were performed where each intensity value was divided by the median of all the values on the chip. The resulting expression levels on each chip are centered around one. Hierarchical clustering was performed using normalized values. Before clustering, the expression values for a gene across all the samples were standardized (linearly scaled) to have a mean of 0 and standard deviation of 1, and the standard values were used to calculate correlation between genes and samples and serve as the basis for merging nodes. The dendrograph shows the clustering of 43 genes and ESTs that were selected for representation. Gene-expression profiles are shown in rows. Red indicates that the gene is expressed more (two and upfold) as compared with basal levels that are shown in green. The rust-color bar indicates relative changes (fold) of mean normalized intensity of normal and OA-affected patients.

EST: Expressed sequence tag; OA: Osteoarthritis; PAC: P1-derived artificial chromosome.

Taken from [9].

Figure 2. Genes modulated in peripheral blood mononuclear cells of active/remission rheumatoid arthritis and osteoarthritis patients.



Total RNA isolated from peripheral blood mononuclear cells (PBMCs) from normal, RA-A, RA-I and OA patients (as per ACR criteria) were used for gene chip analysis screening 35,000 mRNA transcripts. These samples were hybridized against human Affymetrix GeneChip® and further analysis by GeneSpring Software 5.1. The mRNA transcripts were filtered using fold-change. Transcripts that were differentially expressed, at least in one condition, by twofold and greater were selected as shown in this figure. The data demonstrated specific modulation in OA and RA (RA-A and/or RA-I) PBMCs. These studies demonstrated a specific 'transcriptome-based biomarker signature' in human PBMCs that were distinct in normal, OA and/or subsets or RA. These data also demonstrate the genetic complexity associated with these diseases.
OA: Osteoarthritis; RA: Rheumatoid arthritis; RA-A: RA-active; RA-I: RA-inactive.
Taken from [8].

Conclusion

These recent applications of genomics in areas of biomarkers for predictive medicine and diagnosis have changed the tide towards mining meaningful biomarkers for diagnosis, while also combining variable components that play a critical role in the etiology of the disease, thus placing idiopathic OA and RA into a common, multifactorial class of genetic diseases. The coming years will refine the various subsets of OA and RA based on the clinical phenotypes described above and the gene-expression array patterns. Thus, a cornerstone for predictive and personalized medicine will emerge in the coming 5–10 years of clinical diagnosis of arthritis. The integration of clinical data in a hospital system by software such as EPIC will augment the quality of tissue banks for such studies. These initiates have been supported by the US FDA with recommended guidelines [14]. Multiplex real time PCR may impose the practical realities for diagnostic biomarkers in complex diseases, including arthritis. In conclusion, the genomic complexity of the disease and its processes entails precise phenotypic characterization of clinical indicators and application of an array of complex technological approaches to generate meaningful diagnostic biomarkers.

Acknowledgement

We would like to thank Dr Palejwala for editing the manuscript and Mrs Michelle Rothrock for typing the manuscript. Seth Thompson and Shailey Amin were summer students who participated in preparation of the manuscript.

Financial disclosure

The authors have no relevant financial interests including employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties related to this manuscript.

Executive summary

Pursuit of the arthritis gene(s)

- Arthritis is a complex disease involving numerous genes representing various traits associated with the disease process.
- Most of the candidate genes and their products identified to date via an *a priori* hypothesis represent aftermaths (mostly irreversible) of the disease, thus proving inadequate as potential therapeutic targets and/or realistic predictive biomarkers.

Connecting the genetic traits

- Less than 10% of 'arthritis genes' segregate via a Mendelian trait, the rest remain non-Mendelian, segregating via a combination and/or under the influence of other genetic processes.
- Various forms of arthritis with distinct phenotypic characteristics (e.g., hand, knee and/or hip osteoarthritis [OA]) are represented by different clusters of overlapping and/or distinct genes, such as IL-1.
- New approaches for scrutinizing the genome are allowing us to connect the dots in these complex genetic processes.

The nongenetic component that modulates arthritis

- Identical DNA sequences may be silent in one individual, but may be activated and/or modulated in another individual due to aging, environmental factors and lifestyles, all of which impact the progression of arthritis.

Emerging technologies

- Technologies are facilitating the translation of complex phenotypic characteristics of OA and rheumatoid arthritis (and subcategories) directly into specific patterns that represent not only the clinical and phenotypic aspects of arthritis, but also haul out the central players, peripheral components and outliers from a single picture of the disease process.

Conclusion

- Gene expression arrays, multiplex real time PCR and/or total genome walking will continue to dominate identification of novel biomarkers for diagnosis, prognosis and potential therapeutic targets, provided the starting clinical materials are well curated with reliable tissue banks and electronic clinical databases that are annotated specifically for research.
- These technologies and approaches are poised to be used by the US FDA for future clinical trials.

Bibliography

1. Loughlin J: Genetics of osteoarthritis and potential of drug development. *Curr. Opin. Pharamcol.* 3(3), 295–299 (2003).
2. Ikegawa S, Toshiyuki I, Mabuchi A: Genetic analysis of osteoarthritis: toward identification of its susceptibility genes. *J. Orthop. Sci.* 8(5), 737–739 (2004).
3. John S, Eyre S, Myerscough A *et al.*: Linkage and association analysis of candidate genes in rheumatoid arthritis. *J. Rheumatol.* 28(8), 1752–1755 (2001).
4. Newton JL, Harney SMJ, Wordsworth BP, Brown MA: A review of the MHC genetics of rheumatoid arthritis. *Genes Immun.* 5(3), 151–157 (2004).
5. ENCODE Project Consortium: identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146), 799–816 (2007).
6. Lin H, Baron R, Clement-Lacroix P: Aberrant epigenetics in rheumatoid arthritis and osteoarthritis. *Future Rheum.* 2(3), 257–260 (2007).
7. Jüngel A, Baresova V, Ospelt C *et al.*: Trichostatin A sensitizes rheumatoid arthritis synovial fibroblasts for TRAIL-induced apoptosis. *Ann. Rheum. Dis.* 65(7), 910–912 (2005).
8. Amin AR: Genetics and genomics in human osteoarthritis. In: *Osteoarticular Aging: Osteoarthritis Degeneration.* Maitiza Q (Ed.), (2006)./Amin AR: Genética y genoma de la osteoartritis (OA) humana. In: *ARTROSIS.* Quintero M (Ed.), Caracas, Venezuela, 159–164 (2005).
9. Attur MG, Dave MN, Tsunoyama K *et al.*: 'A system biology' approach to bioinformatics and functional genomics in complex human diseases: arthritis. *Curr. Issues Mol. Biol.* 4, 129–146 (2002).
10. Attur MG, Dave M, Akamatsu M, Katoh M, Amin AR: Osteoarthritis or osteoarthrosis: the definition of inflammation becomes a semantic issue in the genomic era of molecular medicine. *Osteoarthritis Cartilage* 10(1), 1–4 (2002).
11. Patel IR, Attur MG, Patel RN *et al.*: TNF-α convertase enzyme from human arthritis-affected cartilage: isolation of cDNA by differential display, expression of the active enzyme, and regulation of TNF-α. *J. Immunol.* 160(9), 4570–4579 (1998).
12. Olsen N, Sokka T, Seehorn CL *et al.*: A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. *Ann. Rheum. Dis.* 63(11), 1387–1392 (2004).
13. van der Pouw Kraan TCTM, Wijbrandts CA, van Baarsen LGM *et al.*: Rheumatoid Arthritis subtypes identified by genomic profiling of peripheral blood cells: Assignment of a type I interferon signature in a subpopulation of patients. *Ann. Rheum. Dis.* 66, 1008–1014 (2007).
14. Frueh FW: Impact of microarray data quality on genomic data submissions to the FDA. *Nature Biotechnol.* 24(9), 1105–1107 (2006).

Affiliations

- *Ashok R Amin*
  Carilion Clinic, 101 Elm Avenue, Research Department, Roanoke, VA 24013, USA
  Tel.: +1 540 985 8499;
  Fax: +1 540 344 7278;
  aramin@carilion.com
- *Seth D Thompson*
  James Madison University, Department of Biology, Harrisonburg, VA 22807, USA
  Tel.: +1 540 330 6957;
  Fax: +1 540 344 7278;
  thompssd@jmu.edu
- *Shailey A Amin*
  Blacksburg High School, 520 Patrick Henry Drive, Blacksburg, VA 24060, USA
  Tel.: +1 540 951 5706;
  Fax: +1 540 951 5714