

# Chapter 2

## Using Functional Genomics to Identify Drug Targets: A Dupuytren's Disease Example

Mirela Sedic, Sandra Kraljevic Pavelic, and Karlo Hock

### Abstract

Research into the molecular mechanism of Dupuytren's disease (DD) illustrates all the problems common to drug discovery in orphan diseases, but also in more commonly investigated ailments. Current findings characterize DD as a disease with complex molecular pathology, with changes in expression of multiple genes and proteins as well as many contributing risk factors. Some of the observed changes include genes and proteins that have been identified in a number of other pathological processes, such as TGF- $\beta$ , some which may be more specific to DD, such as ADAM12, and undoubtedly also some that have yet to be discovered in future studies. When all these results are taken into consideration, it can be deduced that DD is an end result of several pathological processes that can have many points of origin, and probably involves several subtypes that give rise to sufficiently similar clinical symptoms to be unified under a single medical term. Such breadth of view has become possible with the advent of functional genomics methods and system-wide overview of the molecular processes, which highlight molecular players and processes that might not be intuitively obvious from symptoms, as is the case with the observed parallels with wound-healing processes. As functional genomics methods allow researchers to compile a more complete image of the molecular mechanisms involved in DD pathogenesis, they also help to propose new drug targets that can be employed to develop an effective pharmacological treatment for DD. Identification of key molecular players in DD has already benefited from the integration of functional genomics and biocomputational methods, and such approach may reveal new ways how we can interfere with the emergence of the DD phenotype.

**Key words:** Functional genomics, Drug discovery, Dupuytren's disease, Orphan disease, Transcriptomics, Proteomics, Molecular pathway, Bioinformatics

---

### 1. Introduction

#### **1.1. Functional Genomics and Drug Discovery: A General Overview of Research Approaches**

Functional genomics refers to a set of methods that attempt to identify and characterize the functions of genes in an organism. Rather than mapping the genome sequence of the investigated cells, functional genomics aim to understand the relationship between the genome and the phenotype. It focuses on genes that

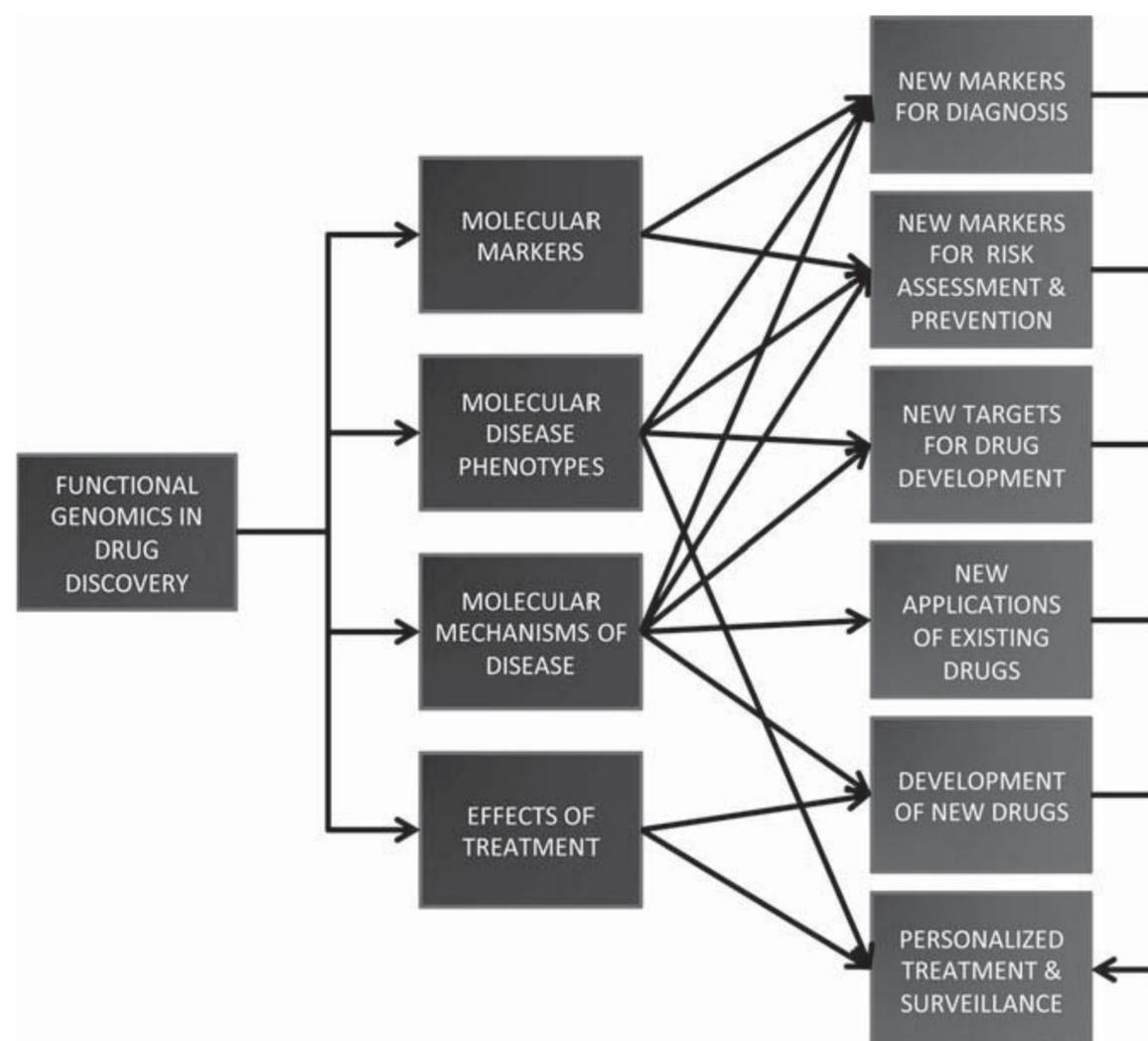


Fig. 1. Potential directions of functional genomics methods in drug discovery.

are actually expressed at a given point in time and their dynamic functions and interactions, thus creating genome-wide profile through which researchers can identify the activity levels of genes in a cell. Since these investigations often involve system-level studies of cellular mechanisms, methods in functional genomics are frequently designed to study the entire macromolecular content of a cell, such as transcriptomic methods for screening the full quantity of messenger RNAs (mRNAs) in a cell or proteomic methods for screening the full quantity of proteins in a cell. In terms of drug discovery, application of functional genomics methods to drug discovery progresses in several research directions (Fig. 1). Functional genomics has generally been used with two principal aims in mind: to elucidate the mechanism of molecular function of existing or potential molecular targets, which then aids in screening of the compounds that can act on such targets, or to predict and propose novel molecular targets for which therapeutics can then be developed. In addition to the main objectives of providing an effective diagnostic markers or treatment for the ailment under investigation, these two directions share the common aims of unraveling the molecular mechanisms that underlie pathological processes and the effects of potential therapeutics on those mechanisms. Functional genomics should therefore aid the transition of current

healthcare system into one that is oriented much more towards prevention and early diagnosis, rather than the one exclusively focused on diagnosis, treatment, and rehabilitation after the onset of disease begun.

Given the advantages in scope and sensitivity of functional genomics methods, this approach lends itself well to the attempts to find the early diagnosis options or effective treatment for the so-called orphan diseases—diseases that have low incidence in a population or for some other reason do not garner as much scientific, clinical, and pharmacological attention as their more thoroughly investigated counterparts. The potential ability of functional genomics methods to pinpoint the differences in molecular processes characteristic of pathological states might make them an effective and efficient way to propose and accelerate the development of diagnostic or treatment options for orphan diseases. While these advantages of functional genomics in drug discovery are equally attractive in research focusing on more common ailments, it is nevertheless a very important aspect of biomedical investigations of orphan diseases given the constrained funding conditions that often accompany such efforts.

**1.2. Functional  
Genomics Helps  
to Identify  
Pathological Molecular  
Processes as well as  
Novel Drug Targets**

Despite the advancement of the new methods developed to identify new molecular targets for drugs, the rate at which new drugs enter the market has not substantially accelerated in the previous decades and has largely remained constant over that period (1). This is in part due to the tightening of regulations when it comes to drug safety concerns. Extensive toxicology studies required prior to clinical testing on humans have resulted in many drug development projects being terminated even in the late stages of research once it has been determined that the compounds might be unsafe for use. Given the length, costs, and uncertainty involved in drug discovery processes, it is of great interest to find ways how the whole process can be made more predictable in terms of the eventual outcome.

Detailed knowledge of the molecular mechanism of action is one of the best ways how this aim can be achieved. While a drug can be therapeutically effective even without intricate insight into its exact mechanism of action, as proven by many famous examples in the history of pharmacology, possessing such knowledge can greatly accelerate the discovery of effective treatment options and at the same time reduce the costs of such process. With their ability to comprehensively study many cellular processes at once, functional genomics methods are well suited for studying molecular mechanism of disease. Rather than giving a static overview of the genetic material available to the cell, for example due to sequencing, functional genomics focuses on dynamic expression profiles of those genes. Methods used in functional genomics usually survey either cell's transcriptome, all mRNA molecules

expressed in a cell, or its proteome, that is, all proteins currently present in a cell. An overview of the cell's mRNA content provides insights into the amount of available mRNA for protein production, and can be correlated with the level of activity of genes coding for it, as well as the speed of mRNA degradation. Transcriptomics methods, such as DNA chips or microarrays, can survey thousands of transcripts simultaneously by reversely transcribing the isolated mRNA and hybridizing it against an array of complementary DNA fragments. Using fluorescent dyes it is then possible to determine not only the presence of specific mRNA molecules with known sequence in the samples, but also the amount of each mRNA molecule that was present in the samples. What is more, it is possible to compare such samples with reference levels of expression, for example from appropriate controls, and therefore obtain a differential expression profile characteristic of a pathological process under investigation. These differences then accentuate either the altered activity of gene transcription for those genes suspected to play a role in the disease pathogenesis, or highlight novel molecular players and potential targets for future research. As such, transcriptomics studies can offer useful insights on whether a pathological process stems from the abnormal transcription, or regulation of transcription, of a gene.

Transcriptomics methods can also be used to simulate the effects of therapeutics aimed at removing a particular gene product from the cell, and therefore to gain insight into whether such option would be a viable path to take in drug development. Eliminating or attenuating the production of mRNA molecules removes the template for production of proteins in a cell (2). A traditional way to manipulate the production of mRNA molecules in a cell is to eliminate it entirely, by using the so-called "knockout" cells or organisms that entirely lack, or for some other reason cannot produce, protein under investigation. However, while such methods provide some insights into the functioning of a cell that lacks a target gene product and consequently into the potential role of a gene product in the pathogenesis, producing genetic knockouts comes with some distinct disadvantages. Methodologically, producing knockouts is often an impractical thing to do and has only a limited use in human studies through in vitro methods. Moreover, knockout mutations often severely impair development and many are inherently lethal, making it impossible to produce a viable knockout model for many diseases. As a consequence, much attention has been devoted to development of techniques to attenuate, rather than completely eliminate, the production of mRNA molecules in cells (2, 3). This can be achieved at several levels of mRNA metabolism, starting with the regulation of mRNA transcription. As transcription of mRNA requires bonding of specific transcription factors around the coding regions of DNA, interference with

these processes, such as through modification of transcription factors or modification of transcription factor binding sites on DNA (for example, through DNA methylation), can affect production of mRNA and consequently the subsequent abundance of the corresponding protein product. The development of methods jointly known as RNA interference (RNAi) has garnered an especially large amount of research attention. The RNAi mechanism is a natural process employed by eukaryotic cells to inhibit protein production at a posttranscriptional level through the use of small strands of RNA that specifically bind to complementary sequences of target mRNA molecules to control their activity, mostly by temporarily eliminating the translation of such double-stranded mRNA into protein. Given that, in contrast with eliminating the gene at the DNA level, the reduction of mRNA levels through these techniques is often temporary and can also be incomplete, these techniques are said to produce a “knockdown” rather than a “knockout” effect in a cell. As these methods tend to be highly successful for gene silencing in experimental settings, they are becoming an indispensable tool for large-scale functional genomics screens and high-throughput drug target screening.

However, while the absence of mRNA molecules can cause the drop in levels of proteins in a cell, the availability of mRNA in a cell does not necessarily translate into the comparable biological activity of proteins (4). In addition to many posttranscriptional control points that regulate translation of mRNA into proteins as well as diverse posttranslational modifications, proteins also form interdependent and overlapping signaling pathways that feature many feedback and feed-forward loops both within and across pathways (5). The behavior of such complex biological networks cannot be predicted from the transcriptome alone, necessitating the development and application of techniques that can provide direct data on the activity of proteins in a cell (6, 7). Proteomics, a joint term used for the methods that attempt to study the protein content of a cell (a proteome), is therefore an important aspect of functional genomics research aimed to uncover the levels of cell metabolic activity, and therefore also the corresponding drug discovery research. Comparisons between affected and control cells, or between control and treated cells, allow for direct links to be made between such parameters and molecular responses in a cell. Therefore, methods that measure the quantitative levels of proteins can be used not only to study the differential expression of proteins resulting from pathological processes, but also the actual effects that stem from the application of therapeutic compounds. Thus, the same set of methods can be used to characterize disease pathogenesis, discover novel drug targets, and monitor the molecular changes that result from drug therapy. As such, large-scale investigations of the proteome, especially comparative studies of proteins expressed in pathologically modified cells with

and without pharmacological treatment, show great promise for identifying potential drug targets and molecular consequences of therapy.

Due to the complexity of functional genomics results, the information gained through functional genomics experiments is often complemented by biocomputing analyses. In general, biocomputing has three major roles (8): (1) to organize data in such a way as to allow researchers to access existing information and to submit new data, e.g., Protein Data Bank for 3D macromolecular structures, (2) to develop tools and resources that would help in data analysis, e.g., comparison of protein sequences, and (3) to use these tools for data interpretation in a biologically meaningful manner. Such analyses are necessary when dealing with the kinds of datasets produced by functional genomics methods, and ultimately serve the purpose of highlighting the results of interest, from macromolecules with differential expression to homologies across systems and integration of system-level data. While not unique to drug discovery research, analyses of this kind can also pinpoint potential therapeutic targets and/or markers needed for early diagnostics.

---

## **2. Molecular Mechanisms, Functional Genomics, and Drug Discovery Perspectives in Dupuytren's Disease**

As a case study illustrating the use of functional genomics in drug discovery, especially in orphan diseases, in the rest of this chapter we provide an in-depth look of the research employing functional genomics methods and biocomputing to find the molecular causes of pathogenesis, and consequently potential drug targets, in Dupuytren's disease (DD). DD is a benign palmar fibromatosis with high prevalence in some populations which also severely impairs the hand function of the affected patients. Despite the frequency of its occurrence, DD has received limited research attention to date. However, the picture that has emerged from the studies to date is a complex one, and DD pathogenesis likely involves several molecular pathways and thus a broad array of potential targets for drug development and early diagnostic markers. As such, it provides a good example of challenges often faced by researchers attempting to unravel molecular mechanism of orphan diseases and find potential drug discovery for such ailments. While a helpful nonsurgical treatment for symptoms of the DD has recently become available (9), the research into the molecular mechanisms that lead to its emergence is still ongoing, as this treatment has not been equally efficient in symptom alleviation in all patients. Molecular players highlighted by functional genomics methods may hold the key to find suitable drug targets to overcome this problem.

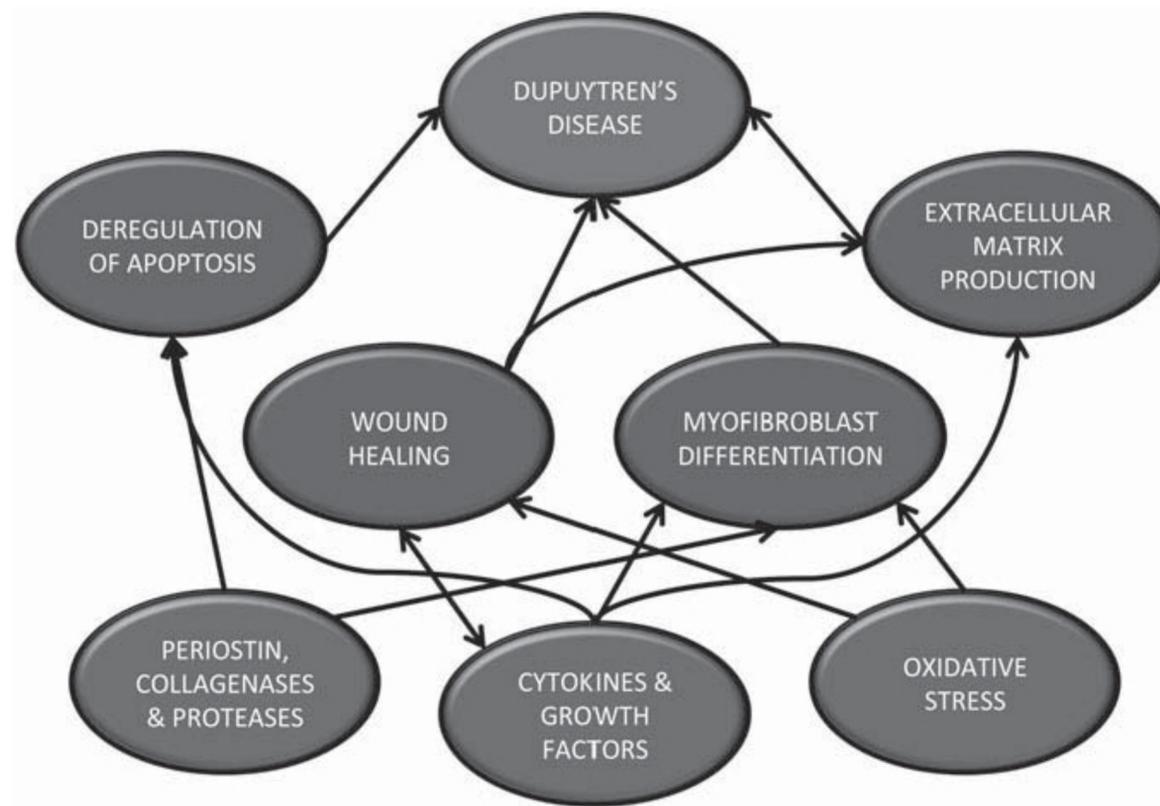


Fig. 2. An overview of major processes that lead to the emergence of Dupuytren's disease symptoms.

### 2.1. Molecular Factors in the Pathogenesis of Dupuytren's Disease

Molecular profile of pathological changes in DD is complex, and involves many molecular factors that belong to several different cellular processes and pathways (Fig. 2). Abnormal growth factor expression, in particular those of cytokines basic fibroblast growth factor (bFGF) and transforming growth factor beta (TGF- $\beta$ ), are believed to primarily drive at least two molecular processes in Dupuytren's disease: (a) proliferation of fibroblasts and their differentiation into myofibroblasts and (b) production of dense extracellular matrix containing elevated levels of fibronectin, type III collagen, and proteoglycans (10–12). Badalamente et al. (13) studied intracellular and extracellular localization of two common TGF- $\beta$  isoforms, namely TGF- $\beta$ 1 and TGF- $\beta$ 2 in the proliferative, involutinal, and residual stages of Dupuytren's disease, and found that TGF- $\beta$ 1 was active in fibroblasts and myofibroblasts at all disease stages. TGF- $\beta$ 2 was localized in myofibroblasts in the proliferative and involutinal stages, but was absent in the residual phase fibroblasts. Importantly, the studies on explant cultures from Dupuytren's nodules in the proliferative or involutinal stage revealed that, compared to control myofibroblasts, the addition of TGF- $\beta$ 1, TGF- $\beta$ 2 and the combination of both isoforms had significant effects on myofibroblast proliferation (13). However, TGF- $\beta$ 2 had the most significant proliferative effect.

There is an evidence that the expression of unstable form of zinc-finger protein 9 (Zf9) could predispose patients for development of Dupuytren's disease (14). Zf9 is a transcription factor synthesized in the cell nucleus, which binds to the promoter region in the genes coding for TGF- $\beta$ 1, TGF- $\beta$ 2, and their

respective receptors. The gene for the Zf9 protein containing the nucleotide guanine at position 1,140 synthesizes an unstable form of this protein, which appears to be directly responsible for increased synthesis of TGF- $\beta$ 1, TGF- $\beta$ 2, and their respective receptors in serum and tissue (15).

Several studies confirmed that structural components of the extracellular matrix such as collagens (particularly collagen type III), laminin, fibronectin, and elastin are altered in DD (16, 17). After quick synthesis of immature, normal-length collagen type III, contractile force of myofibroblasts causes collagen structure of tissues to shorten, leading to an increase of contractile force and loss of hand function. This process is believed to be the result of several molecular events: increased density of fibroblasts, stimulation of fibroblasts by growth factors, decreased rate of apoptosis (programmed cell death) of fibroblasts, and disproportion between collagenases (also known as matrix metalloproteinases, MMPs) and their inhibitors, also known as tissue inhibitors of metalloproteinases or TIMPs (10). Recent work has also identified proteases, including A disintegrin and metalloprotease (ADAM)-12, proteoglycans (notably PRG4), and “matricellular” components, including tenascin C and periostin, as well as specific members of the metalloprotease family (MMP-2 and MMP-9) as being abnormally regulated in DD (18–20). Importantly, Vi et al. (20) showed that periostin is abundant in Dupuytren’s disease cord tissue. The relevance of periostin upregulation in DD was assessed in primary cultures of cells derived from diseased and phenotypically unaffected palmar fascia from the same patients. These studies revealed that periostin induces the proliferation and apoptosis of phenotypically normal fibroblasts derived from the palmar fascia adjacent to the DD cord, and myofibroblast differentiation of DD cells. Altogether, it seems possible that periostin, secreted by disease cord myofibroblasts into the extracellular matrix, promotes the transition of resident fibroblasts in the palmar fascia toward a myofibroblast phenotype, thereby promoting disease progression (20).

It has also been hypothesized that immunological response driven by several cytokines plays an important role in the pathogenesis of Dupuytren’s disease. For example, IL-1 that stimulates platelets and macrophages giving rise to the secretion of several growth factors (e.g., TGF- $\beta$ , FGF (fibroblast growth factor) and PDGF (platelet-derived growth factor)) has been shown to induce migration of inflammatory cells to the fascia affected by DD (21). The latter cells produce different growth factors, including TGF- $\beta$  (22), leading ultimately to palm contracture. In turn, increased secretion of IL-1, TGF- $\beta$  and fibronectin in DD accounts for rise in the number of inflammatory cells, and consequently, further increase in the production of cytokines and growth factors by these cells. Several authors have hypothesized that this phenomenon occurs as a consequence of completely abrogated/partially

deregulated apoptosis either in fibroblast/myofibroblast cells or in inflammatory cells in DD tissues (23, 24).

Strong evidence exists on the possible role of oxidative stress in DD pathogenesis, showing an increased release rate of superoxide free radicals ( $O_2^-$ ), hydrogen peroxide ( $H_2O_2$ ) and hydroxyl radicals ( $OH^-$ ) in affected palmar fasciae (25). Oxygen free radicals at lower concentrations, i.e., concentrations similar to those likely to occur in Dupuytren's disease, might stimulate proliferation of fibroblasts in DD (25). Such increased proliferation rate of fibroblasts may lead to accumulation of type III collagen and fibronectin. In the wound, the synthesis of both collagen II and fibronectin exceeds their degradation rates, followed by rise in the collagen levels. Later on, in normal tissue, production and degradation of collagen become equal. On contrary, in DD, a net collagen gain might be observed in affected fasciae. It might not be ruled out that high levels of free radicals in DD tissues might be produced by defective mitochondria, in which free radicals are generated by means of electron leak, as it was reported that a mutation within the mitochondrial genome (16S rRNA region) in patients with Dupuytren's disease exists (26).

**2.2. New Insights  
into Dupuytren's  
Disease Pathogenesis  
Revealed by High-  
Throughput Functional  
Genomics Methods**

Until last decade, most issues in modern cell biology have been tackled using the so-called "reductionist methods," i.e., by studying one gene, one protein, one specific protein modification, or one molecular event at a time. This reductionism has been justified, given the complexity of biological systems and lack of appropriate tools for developing more integrative methodologies at that time. However, with a perpetual technological advancement, it has now become possible to get a more thorough understanding of complex biological systems by observing their features and dynamics at different molecular levels. This integrative functional genomics approach makes it possible to reveal intricate networks of molecular interactions and cellular pathways by using high-throughput and large-scale methodologies, such as transcriptomics and proteomics, combined with statistical and computational analyses of obtained results (27). These cellular interaction maps, with assigned dimensions of time and space, are the answer to challenges faced by molecular complexity, and ultimately reveal the big picture of cellular functions in disease states. Such integrated networks are particularly pertinent to complex diseases, where targeted therapy against single proteins is insufficient, and more effective therapeutic approach can be applied only with complete comprehension of integrated molecular profiles.

Transcriptomic studies of DD unveiled specific "gene signatures" pointing to cellular events that might account for DD pathogenesis, and which include collagen degradation, generation of the contractile force, myofibroblasts differentiation, oxidative stress, regulation of apoptosis, proteolysis and inflammation, fibrosis, cytoskeletal development, lipid metabolism, cell growth,

proliferation, differentiation, regulation of cell death, biological cell adhesion, localization, extracellular matrix–receptor interaction, cell communication, and ossification. These molecular findings were not entirely unexpected, as all of them can be linked with reported DD symptoms. However, several novel findings came into focus revealing that DD shares some common gene expression patterns with liver fibrosis, e.g., an overexpression of collagen type V  $\alpha$ 2 (COL5A2), ADAM metalloproteinase domain 12 (ADAM12), and cysteine and glycine-rich protein 2 (CSRP2), along with downregulation of procollagen C-endopeptidase enhancer 2 (PCOLCE2) and matrix metalloproteinase 3 (MMP3) (18). Importantly, observed difference in the expression status between the major deregulated genes from the nodules and from the cords in comparison with external control fascia clearly underlines the potential of transcriptomic profiling for discerning DD phenotype.

MicroRNAs (miRNAs) are classes of small noncoding RNAs (~22 nt) which normally function as negative regulators of target mRNA expression at the posttranscriptional level (28). Their binding to the 3'UTR of target mRNAs through base pairing brings about the target mRNAs cleavage or translation inhibition. Their roles in many crucial biological processes have been ascertained including cell growth, tissue differentiation, cell proliferation, embryonic development, and apoptosis. Consequently, mutation of miRNAs, dysfunction of miRNA biogenesis, and dysregulation of miRNAs and their targets might result in various diseases. Mosakhani et al. (29) were the first to characterize the miRNA profiles of DD patients, and found that some of the identified miRNAs regulate the genes related to the  $\beta$ -catenin pathway, namely WNT5A, ZIC1, and TGFBI. This study revealed deregulation of  $\beta$ -catenin pathway in DD and proposed the miRNA expression profiling as a promising tool for novel research strategy for DD.

As proteins are regarded as functional outputs of the cells, studies of their biological activity, expression, structure, and function are very important in terms of disease pathogenesis. Proteomics methods help to advance our understanding of cellular behavior at the systems level, leading to identification of new drug targets and the development of novel diagnostic tests. In addition, proteomics can reveal the alterations that RNA analysis alone cannot provide, such as protein–protein interactions, cellular localization, modifications (such as isoforms and posttranslational modifications), and protein abundance. However, study of proteins is hampered by the possible involvement of many different cellular regulatory mechanisms that ultimately shape their structure, function, and expression. Unlike genomes, proteomes are dynamic, and many proteomic studies focus on examining changes in proteome composition under various conditions (e.g., diverse stages of disease progression). Two-dimensional gel electrophoresis (2DE) is a

powerful method for separation, visualization and quantification of thousands of different protein species in a single run. Differentially expressed proteins resolved on 2DE gels can then be identified by two different approaches, either by peptide-mass mapping using MALDI-TOF MS (Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry), or by peptide sequencing using ESI-MS/MS (electrospray ionization-tandem mass spectrometry). Using 2DE/MALDI-TOF MS, a comparative study between affected and adjacent, nonaffected patient-matched tissues from patients with DD in the involitional phase has been carried out by Kraljevic Pavelic et al. (30). The results highlighted alterations in the proteins associated with extracellular matrix (ECM) production, cell proliferation and differentiation, cytoskeleton assembly and maintenance, muscle contraction, energy production (glycolysis and citric acid cycle), regulation of apoptosis, and response to oxidative stress. Moreover, these proteins were identified in the context of their respective biological pathways by creating the protein-protein interaction network (interactome). Such approach resulted in the discovery of several signaling events novel to the Dupuytren's disease, namely auto-crine regulation through ERBB-2 and IGF-1R receptors, and Akt signaling pathway in prosurvival signaling in DD fibroblasts. These results demonstrate that proteomics complemented with bio-computing tools for data processing might be an effective paradigm to reveal new molecular processes in disease pathogenesis, which will set new directions in DD research (30). The increasing use of high-throughput platforms for the analysis of protein expression levels driven by technological advancement in mass spectrometry and array-based technologies has emerged as potential diagnostic tool for rapid and accurate high-throughput screening. SELDI-TOF MS (Surface Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry) couples array-based technology (ProteinChip<sup>®</sup>, Ciphergen Biosystems Inc.) with MALDI-TOF MS. Protein mixture is applied onto the surface of the protein chip arrays that retain the proteins according to their specific physico-chemical properties. Comparisons of spectra obtained from large number of different samples reveal unique or overexpressed protein signal in a particular sample set. Using this approach, O'Gorman et al. (31) analyzed normal and disease palmar fascia from DD patients, and detected 14 upregulated and 3 downregulated low molecular weight (2–20 kDa) peptides and/or proteins. Further bioinformatics processing of obtained data confirmed three disease-associated protein species (4,600.8 Da, 10,254.5 Da, and 11,405.1 Da) that were elevated (5.45, 11.7, and 4.28-fold, respectively). ProteinChip<sup>®</sup> Technology might potentially help researches to focus on specific molecular aberrations in diseases with unknown etiology, such as DD, and to lay the groundwork for discovering new therapeutic and/or diagnostic targets.

**2.3. Biocomputational  
Analysis Complements  
Functional Genomics  
Methods and Aids in  
Drug Target  
Identification**

While the results from both transcriptomics and proteomics methods can provide useful insights into the molecular mechanisms of pathological processes or treatment effects, functional genomics results of DD investigations accentuate the fact that the observed expression profiles alone often do not tell the complete story of disease pathogenesis. Cells as biological systems feature a large degree of redundancy, with pathways that are interconnected at many levels (5, 6, 32). This intrinsic complexity of biological systems means that molecular pathways are far from being independent of each other and in fact exist as a complex web of interdependent components and subsystems. In terms of interpreting the results of high-throughput methods, this means that detecting differential expression of macromolecules (proteins as well as mRNA) does not necessarily imply that those particular molecules are in fact responsible for the observed pathological processes. Indeed, such high expression levels might be indicative of a mechanism acting to compensate for the original source of the disease (32). Moreover, functional genomics results often contain hundreds or even thousands of identified molecules, and need to be further analyzed and contextually interpreted to extract the important information. Biocomputational methods provide this crucial final step in the biological interpretation of the functional genomics results and, consequently, in identification of potential drug targets for future research.

The first step in biocomputational analysis is often to standardize the terms used in the experimental dataset. Functional genomics datasets that emerge from different identification techniques, such as microarrays or mass spectrometry, often also incorporate different terminology standards, especially if both transcriptomics and proteomics methods have been performed at different stages of investigation. It is therefore necessary to ensure that the terminology used will be applied uniformly and suitable for in-depth analysis. Several online tools can be very useful in this regard, as they contain the information needed for successful conversion between different encoding formats. For example, various databases maintained by National Center for Biotechnology Information, freely available at <http://www.ncbi.nlm.nih.gov>, Swiss Institute for Bioinformatics (33), freely available at [www.expasy.org](http://www.expasy.org), or Gene Ontology Project (34), freely available at <http://www.geneontology.org>, contain the information which can aid in linking identified proteins to the genes that code for them, and therefore allow for easier comparisons between transcriptomic and proteomic data. Tools of this kind are also useful for finding synonyms, as the results obtained by functional genomics methods may also use different nomenclature from the computational tools used later in the analysis process. Moreover, such tools can also help when it comes to hypothetical proteins

that sometimes form a part of the experimental dataset, as well as to resolve dilemmas in situations when, for example, more than one target protein can be the result of the spectrometric analysis.

Once the data has been converted into a desired format, other biocomputational tools can aid in biological interpretation of functional genomics results. A very useful step is the categorization of identified genes or gene products in accordance with their function and cellular localization, as well as pathways and biological processes that they belong to. Analyzing the data with tools that can provide such answers greatly aids in the interpretation of the results, as it becomes possible to discuss the results in terms of interdependent processes rather than single molecules. This facilitates systems thinking and biological interpretation, as it is often more convenient to envision how the interaction and interdependencies between several subsystems, instead of hundreds of individual elements, give rise to the emergence of disease pathogenesis. Tools like PANTHER Classification System (35), freely available at <http://www.pantherdb.org>, can greatly aid in this effort, as inputting a simple list of macromolecules returns the information about the involved pathways, as well as homologies between various biological and pathological processes. Moreover, identified genes or gene products can also be classified according to the control of their regulation. Coregulation analysis, determining whether the identified genes could be regulated by common transcription factors, brings the interpretation another step closer to understanding the molecular mechanism of disease, and is equally important in search for potential triggers of pathological processes as it is in considerations whether some differentially expressed genes should be considered as crucial for pathological changes or simple consequences of shared mechanism of transcriptional regulation. Tools that can identify commonalities in transcription factor binding sites, such as Distant Regulatory Elements of coregulated genes or DiRE (36), freely available at <http://dire.dcode.org>, can provide just such insights, and, for example, help determine whether the obtained changes in expression levels of proteins could stem from regulation at the level of transcription or from some other molecular events. This kind of analysis can narrow down the search for suitable drug target candidates, which is especially vital when scientific information about the biological process under investigation is limited.

Understanding of complex datasets, such as those produced by functional genomics methods, is often aided by data visualization. One of the most striking categories of such visualizations are interactomes, or protein–protein interaction maps. Using tools that utilize either curated databases or text mining to find

interactions and connect proteins, it is possible to use a simple list of proteins obtained from the functional genomics methods and put it into a biological context as an interconnected network. One of the freely available tools for interactome generation is Search Tool for the Retrieval of Interacting Genes or STRING (37), <http://string-db.org>, which uses both predicted and experimental interactions between proteins. While in an ideal case one would end up with a single map that contains all of the inputted proteins, the end result is rarely this convenient. Disconnected elements of a network may appear for various reasons, for example because a particular interaction is missing from the database or because some proteins that could otherwise link up the disconnected elements were not detected as being differentially expressed by the experiment. Search for such links creates opportunities for further research, which can include finding novel interactions between proteins or highlighting proteins further up the signaling cascade that could play a role in disease pathogenesis. In the same way as the identification of the differentially expressed proteins, filling in such gaps in the interactome map can not only stimulate further research into the molecular processes of disease pathogenesis but also help to draw attention on potential drug targets for future investigations.

Taken together, computational tools, such as the ones listed here, can aid in our efforts to interpret functional genomics results in a biological context, and as such advance and accentuate the potential of functional genomics as a method to identify drug targets or diagnostic markers. However, computational tools only provide a solution for a single stage in the drug discovery process, and should be followed by experimental validation of selected targets. While they cannot provide a definite answer on the most promising targets for drug development, they can nevertheless greatly aid in narrowing down the pool of potential target candidates and also bring into focus the molecules and processes that have not been previously linked with the disease under investigation. As noted, this combination of functional genomics methods and biocomputational analysis has recently highlighted the role of novel and previously unexplored molecular pathways, such as auto-crine regulation through ERBB-2 IGF-1R receptors and Akt signaling pathway in DD (30, 38). The only currently approved drug for nonsurgical treatment of DD targets collagen, one of the major molecular players contributing to the rise of symptoms, but as such it might not target the root of the problem that would allow for disease prevention or elimination. This emphasizes the need to characterize novel targets in DD through functional genomics and biocomputing methods in order to understand the emergence of the immediate symptoms from a complex biological network of molecular interactions. Given that the discovery of new molecular processes involved in disease pathogenesis also provides potential

targets for therapeutics, such integrative approach holds great promise when it comes to its future applications in therapy and diagnostics for DD.

**2.4. Conclusions  
and Perspectives:  
Functional Genomics  
and Drug Discovery in  
Dupuytren's Disease**

During the last decade, large-scale implementation of functional genomics methods in basic life science and medical research has fostered an exponential growth in genome-wide data collection and measuring. As such, the functional genomics approach is proving to be extremely valuable for disease studies in general, and orphan diseases like DD in particular. The example of advances in our understanding of DD that have been made possible by functional genomics methods illustrates how such methods can be used to shed additional light to the pathogenesis of diseases, and reveal a range of molecular factors contributing both to disease onset as well as the development of disease symptoms. However, it is still difficult to gain a systems-level understanding of a pathological process in DD only from basic interpretation of obtained results, such as lists of genes or proteins. Integration of functional genomics with biocomputing along with other disciplines may help to overcome these obstacles in data interpretation. Indeed, the sheer number of identified drug target candidates from functional genomics studies may emerge as a bottleneck in development of an appropriate drug to treat DD. Current validation methods are, however, often tailored for a particular gene and require individualized, time-consuming, and expensive studies. Here again, the hope is that the technologies that filter the candidates in a high-throughput, cost-effective manner, such as RNAi and proteomics, might represent a contemporary framework for approaching target validation in drug discovery.

The question is whether the application of functional genomics methods in drug discovery process can be tailored so as to provide viable diagnostic or treatment options for a disease as complex as DD. Three general approaches may be employed to tackle this problem: the first is to treat immediate symptoms, as has already been done by Hurst et al. (9); the second would be to try to identify key players that give rise to DD pathology in most, if not all, DD cases; and the third would be to use case-specific diagnostics and personalize the therapy for each patient using the available treatment options. Regardless of the approach of choice, functional genomics methods can be useful tools to achieve this aim, as they not only help to identify targets for intervention into the pathological processes but also provide clues on how the system will change following therapeutic intervention. Using functional genomics in screening of potential drugs and prodrugs may therefore provide key insights into the action and effectiveness of candidate compounds, thereby accelerating drug discovery research in DD and making it more cost-effective. These aspects are important in any pharmacological research, but are absolutely crucial for

orphan diseases, in which feasibility of drug discovery is often limited by the strict economic or experimental constraints. Functional genomics, with its ability to survey a wide spectrum of potential targets, may prove invaluable in this regard. Due to the complex nature of cell's molecular machinery these efforts will likely require ever tighter integration of experimental functional genomics methods with appropriate biocomputing methods. Methods such as interactome analysis may help identify key components that are not immediately obvious from expression results, and as such provide a cost-effective way to predict the functioning of DD pathogenesis and propose the direction of future studies. This combination of functional genomics and biocomputational methods holds great promise as a paradigm for future research in drug discovery in DD and other orphan diseases, as well as in other more commonly investigated ailments.

## References

1. Kramer R, Cohen D (2004) Functional genomics to new drug targets. *Nat Rev Drug Discov* 3:965–972
2. Savoie CJ, Aburatani S, Watanabe S, Eguchi Y, Muta S, Imoto S, Miyano S, Kuhara S, Tashiro K (2003) Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades. *DNA Res* 10:19–25
3. Dorsett Y, Tuschl T (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* 3:318–329
4. Dove A (1999) Proteomics: translating genomics into products? *Nat Biotechnol* 17:233–236
5. Kitano H (2002) Computational systems biology. *Nature* 420:206–210
6. Araujo RP, Liotta LA, Petricoin EF (2007) Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat Rev Drug Discov* 6:871–880
7. Ryan TE, Patterson SD (2002) Proteomics: drug target discovery on an industrial scale. *Trends Biotechnol* 20:S45–51
8. Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med* 40:346–358
9. Hurst LC, Badalamente MA, Hentz VR, Hotchkiss RN, Kaplan FT, Meals RA, Smith TM, Rodzvilla J (2009) Injectable collagenase clostridium histolyticum for Dupuytren's contracture. *N Engl J Med* 361:968–979
10. Al-Qattan MM (2006) Factors in the pathogenesis of Dupuytren's contracture. *J Hand Surg Am* 31:1527–1534
11. Seyhan H, Kopp J, Schultze-Mosgau S, Horch RE (2006) Increased metabolic activity of fibroblasts derived from cords compared with nodule fibroblasts sampling from patients with Dupuytren's contracture. *Plast Reconstr Surg* 117:1248–1252
12. Baird KS, Crossan JF, Ralston SH (1993) Abnormal growth factor and cytokine expression in Dupuytren's contracture. *J Clin Pathol* 46:425–428
13. Badalamente MA, Sampson SP, Hurst LC, Dowd A, Miyasaka K (1996) The role of transforming growth factor beta in Dupuytren's disease. *J Hand Surg Am* 21:210–215
14. Bayat A, Watson JS, Stanley JK, Ferguson MW, Ollier WE (2003) Genetic susceptibility to dupuytren disease: association of Zfp9 transcription factor gene. *Plast Reconstr Surg* 111:2133–2139
15. Cordova A, Tripoli M, Corradino B, Napoli P, Moschella F (2005) Dupuytren's contracture: an update of biomolecular aspects and therapeutic perspectives. *J Hand Surg Br* 30:557–562
16. Howard JC, Varallo VM, Ross DC, Faber KJ, Roth JH, Seney S, Gan BS (2004) Wound healing-associated proteins Hsp47 and fibronectin are elevated in Dupuytren's contracture. *J Surg Res* 117:232–238
17. Howard JC, Varallo VM, Ross DC, Roth JH, Faber KJ, Alman B, Gan BS (2003) Elevated

- levels of beta-catenin and fibronectin in three-dimensional collagen cultures of Dupuytren's disease cells are regulated by tension in vitro. *BMC Musculoskelet Disord* 4:16
18. Rehman S, Salway F, Stanley JK, Ollier WE, Day P, Bayat A (2008) Molecular phenotypic descriptors of Dupuytren's disease defined using informatics analysis of the transcriptome. *J Hand Surg Am* 33:359–372
  19. Shih B, Brown JJ, Armstrong DJ, Lindau T, Bayat A (2009) Differential gene expression analysis of subcutaneous fat, fascia, and skin overlying a Dupuytren's disease nodule in comparison to control tissue. *Hand (NY)* 4:294–301
  20. Vi L, Feng L, Zhu RD, Wu Y, Satish L, Gan BS, O'Gorman DB (2009) Periostin differentially induces proliferation, contraction and apoptosis of primary Dupuytren's disease and adjacent palmar fascia cells. *Exp Cell Res* 315:3574–3586
  21. Qureshi FI, Hornigold R, Spencer JD, Hall SM (2001) Langerhans cells in Dupuytren's contracture. *J Hand Surg Br* 26:362–367
  22. Gudmundsson KG, Arngrimsson R, Arinbjarnarson S, Olafsson A, Jonsson T (1998) T- and B-lymphocyte subsets in patients with Dupuytren's disease. Correlations with disease severity. *J Hand Surg Br* 23:724–727
  23. Jemec B, Grobbelaar AO, Wilson GD, Smith PJ, Sanders R, McGrouther DA (1999) Is Dupuytren's disease caused by an imbalance between proliferation and cell death? *J Hand Surg Br* 24:511–514
  24. Meek RM, McLellan S, Reilly J, Crossan JF (2002) The effect of steroids on Dupuytren's disease: role of programmed cell death. *J Hand Surg Br* 27:270–273
  25. Murrell GA, Francis MJ, Bromley L (1990) Modulation of fibroblast proliferation by oxygen free radicals. *Biochem J* 265:659–665
  26. Bayat A, Walter J, Lambe H, Watson JS, Stanley JK, Marino M, Ferguson MW, Ollier WE (2005) Identification of a novel mitochondrial mutation in Dupuytren's disease using multiplex DHPLC. *Plast Reconstr Surg* 115:134–141
  27. Kraljevic S, Stambrook PJ, Pavelic K (2004) Accelerating drug discovery. *EMBO Rep* 5:837–842
  28. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q (2008) An analysis of human microRNA and disease associations. *PLoS One* 3:e3420
  29. Mosakhani N, Guled M, Lahti L, Borze I, Forsman M, Paakkonen V, Ryhanen J, Knuutila S (2010) Unique microRNA profile in Dupuytren's contracture supports deregulation of beta-catenin pathway. *Mod Pathol* 23:1544–1552
  30. Kraljevic Pavelic S, Sedic M, Hock K, Vucinic S, Jurisic D, Gehrig P, Scott M, Schlapbach R, Cacev T, Kapitanovic S, Pavelic K (2009) An integrated proteomics approach for studying the molecular pathogenesis of Dupuytren's disease. *J Pathol* 217:524–533
  31. O'Gorman D, Howard JC, Varallo VM, Cadieux P, Bowley E, McLean K, Pak BJ, Gan BS (2006) Identification of protein biomarkers in Dupuytren's contracture using surface enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI-TOF-MS). *Clin Invest Med* 29:136–145
  32. Bailey JE (1999) Lessons from metabolic engineering for functional genomics and drug discovery. *Nat Biotechnol* 17:616–618
  33. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784–3788
  34. The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29
  35. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13:2129–2141
  36. Gotea V, Ovcharenko I (2008) DiRE: identifying distant regulatory elements of co-expressed genes. *Nucleic Acids Res* 36:W133–W139
  37. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks P, Stark M, Muller J, Bork P, Jensen LJ, von Mering C (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 39:D561–568
  38. Sedic M, Jurisic D, Stanec Z, Hock K, Pavelic K, Kraljevic Pavelic S (2010) Functional genomics in identification of drug targets in Dupuytren's contracture. *Front Biosci* 15:57–64