ENGINEERING AND ANALYTICAL METHODS FOR PRECISION MEDICINE

by

Elisabetta Sciacca

PhD Program in

Complex Systems for Physical, Socio-economic and Life Sciences

XXXIII Cycle

Catania, Italy

September, 2020

Abstract

Aim of this thesis is to exploit modern technologies and statistical methods to support precision medicine in both the gathering and analysis of data.

Real-time measurements of clinical parameters and identification of molecular signatures are thought to be crucial to tailor medications on the individual needs and likeliness of response.

This thesis explores both aspects. The first part of the thesis studies the issues related to the safe real-time collection of parameters in connection to ultrasound communications of intra-body sensor networks. A testbed has been developed and a complete characterisation of a high water concentration channel — comparable to human body — is provided.

The second part focuses on the analysis of transcriptomic data to predict successful treatments in non-communicable diseases. To reach this aim we will present a novel pathway-based tool together with the results from two clinical trials.

Acknowledgments

This thesis is the result of the work I carried out at the University of Catania in collaboration with the William Harvey Reasearch Instute (Queen Mary University, London) where I spent the third year of my PhD program.

I am greatly thankful to all the people who supported me and gave me the opportunity to walk through this path.

In particular, I would like to thank my supervisors Laura Galluccio, Alfredo Pulvirenti and Alfredo Ferro as part of the Catania team, as well as my British advisors Myles John Lewis and Costantino Pitzalis in London.

Additionally, I would like to thank all the valuable people who contributed to the work described in this thesis. In particular:

- Laura Galluccio for Chapter 2

- Vito Latora for Chapter 3

- Anna Surace for Chapter 4, Section 4.1

- Elena Pontarini for Chapter 4, Section 4.2

Last but not least, I would like to thank my beloved partner Roberto Paqualino who kept me supporting over these years and gave me fundamental

advice to face all the challenges I dealt with.

Equally, I must particularly thank my parents who tirelessly encouraged me.

Table of Contents

List of Tables

List of Figures

xii

Chapter 1

Introduction

In recent years medicine is rapidly moving away from the generalised treatment approach to follow a strategy to tailor medication on the individuals needs and likeliness of response. This concept is called *stratified* or *precision medicine* and takes advantage of key identifiable genomic, molecular and clinical features [1].

The aim of this thesis is to exploits modern technologies and statistical methods to support precision medicine in both the collection and analysis of data. In order to do so, it is fundamental to consider a patient as an individual subject rather than an average, random sample. This necessarily involves real-time measurements of clinical parameters and identification of genomic signatures [2–4] .

This thesis explores the two topics and is organized as follows. The first section (Chapter 2) addresses the issue of collecting clinical data in a safe manner. Safe and reliable real-time monitoring of clinical parameters is crucial to personalise and adjust ongoing therapies. On the other hand, at a previous stage, the choice of the right therapy should not be based on early symptoms only,

but also on molecular signatures able to predict chances of response. This is especially true for those diseases that show high probability of therapy failure. For this reason, the second part of this thesis (Chapters 3 and 4) focuses on methods and analyses of RNA-sequencing data. In Chapter 3 a novel pathway-based tool called Pat-BINE (Patients Biological Interactions Network Extractor) is presented. This can give clinicians a deeper understanding of the complex network of gene interactions activating in whichever user-defined group of patients. Furthermore, for a higher-level analysis, a tool comparing pathway alterations across multiple groups of patients will be discussed in the same Chapter.

In Chapter 4 novel and traditional methods are applied to two clinical trials. One relies on the Pathobiology of Early Arthritis Cohort (PEAC) [5], the other relates to the transcriptomic data of the TRial of Anti-B-Cell Therapy In patients with primary Sjögren's Syndrome (TRACTISS) [6]. From the two cohorts, new insights on rheumatoid arthritis and Siögren's Syndrome are extracted.

Finally, Chapter 5 provides the conclusions and sketches future research directions.

Chapter 2

Real-time monitoring of clinical parameters

The use of ultrasounds for a safe and reliable communication across sensors

The content of this chapter is based on the work published on Computer Networks as [7].

Organs and tissues monitoring is a broadly recognised imperative to allow adjustment of therapies and prevention of serious events such as hearth attacks or ischemias.

To this purpose the use of implanted or ingested pill-sized sensor devices is foreseen. This leads to a number of sensors/actuators deployed inside the body building up a network that periodically communicate to a remote medical center by means of a gateway (e.g. a smart watch or phone).

Apart from issues related to compatibility of materials, a relevant communication problem is posed in terms of waves being used. In fact, the human body is composed for more than 65% by water, a fluid through which traditional

Radio Frequency (RF) waves scarcely propagate. Moreover, RF waves used in wireless transmission are ionizing. This can cause overheating of tissues and consequent degeneration and damage of cells [8]. To this purpose, ultrasonic waves have been proposed as a safe and efficient alternative to RF for transmission of data inside the body.

Studies in [9] and [10] revealed that ultrasonic communications require much lower transmission power with respect to the low-power RF-based technology as water aids the propagation of ultrasonic waves and hinders the RF transmission. This leads to lower energy per bit cost and longer links supported.

On the other hand, it is widely recognized that therapies exploiting ultrasounds are safe and do not exhibit dangerous counter-effects (in [11] a rich overview of therapeutic ultrasounds applications is presented).

However in order to perform efficient ultrasonic communications inside the body, a deep understanding of the peculiar features of this channel is needed. In this chapter, we will present a complete experimental characterization of the impulse response inside a human body channel.

In order to perform such characterization we developed a testbed which employs a human phantom of ballistic gel to mimic the human body propagation features and perform a set of measurements to fully describe the channel behavior as a function of the employed frequency, composition of the medium and distance.

The rest of this chapter is organized as follows. In Section 2.1 relevant literature in the field is discussed. In Section 2.2 some basics of ultrasounds physics are introduced and in Section 2.3 the main features of a ultrasonic transducer

are recalled. In Section 2.4 the impulse response model for the channel under consideration is derived and the measurement procedure carried out is described in Section 2.5. In Section 2.6 numerical results are illustrated and, finally, in Section 2.8, concluding remarks are drawn.

2.1 Related Work

The use of ultrasonic communications for health monitoring and diagnosis has been proposed in the last years by different research groups. Indeed, together with the use of ultrasounds in early diagnosis of fetus during pregnancy, it was also proposed to employ them for communications inside the human body.

In 2008 for the first time Infantis and Kalis [12] proposed a simple propagation model exploring the theoretical feasibility of using ultrasounds inside the body. In the same line of reasoning, in 2012 Santagati et al. published two papers ([13] and [14]) presenting a detailed discussion on important trade-offs regarding the frequency range, the transmission power, the bandwidth and the transducers' features.

Two years later the same authors developed the first simple prototype of a software-defined testbed architecture providing an experimental demonstration of the feasibility of employing ultrasonic communications into human body [15]. In the same year, also an hybrid opto-acoustic communication among nanorobots was proved [16].

These feasibility studies were independently confirmed by Dressler, Charthad

and Singer in 2015 and 2016, respectively. In particular, the first paper addressed the control and communication problem between intrabody nanosensors and an external unit [17]. Charthad, instead, showed the implementation of a mm-sized implantable device using ultrasonic power transfer and a hybrid bi-directional data communication link [18]. Similarly other studies assessing the possibility to use ultrasonic waves inside the human body appeared [19]. Finally Singer [20] showed an experimental acoustic transmission through real pork tissues and beef liver achieving data rates up to 20-30 Mbps with QAM modulation. On the same line of reasoning, in [21] the use of different modulation schemes was considered in order to identify which mechanism performs better in a variable frequency range, up to approximately 400 kHz. In 2017, in [22] and [23] it was demonstrated for the first time also the possibility of using multihop communications with four ultrasonic nodes communicating through a human body phantom containing organic tissues (chicken bone, muscles and skin). In the same year, in [22], the tissue overheating problem is faced for the first time and a thermal aware routing protocol is proposed.

Different research groups have also focused their efforts on identification of communication schemes for this peculiar environment. In 2015, for the first time, Guan et al. considered the problem of designing optimal network control algorithms for distributed networks like BANs. They proposed lightweight, asynchronous, and distributed algorithms for either rate control or stochastic channel access designed to maximize the throughput under energy constraints ([24, 25]). On the other hand, in [26] the authors proposed an Ultrasonic WideBand (UsWB) transmission scheme combined with a multiple access

technique, based on the idea of transmitting information bits spread over very short pulses in a time-hopping pattern.

In [27] and [28] also other communication schemes and error control strategies are discussed.

In [29] it was also proposed to employ ultrasounds at the nanoscale to design a hierarchical Body Area Nano-Network (BANN) architecture consisting of two types of devices, nanonodes and nanorouter which operate in the THz band. In particular a human hand scenario is considered by taking also into account the impact of path loss and molecular absorption noise.

However, in order to proceed with the implementation and development of real acoustic body area networks, a proper channel impulse response analysis is still needed. In [30] for the first time a simple channel model based on simulation and in-vivo experimental measurements is discussed.

Later, in 2016, Rivet et al. developed a near and far field model based on a simulated body channel [9]. More recent works such as [31] and [32], instead presented a more detailed characterization of channels for frequencies up to 2 MHz.

Based on the above considerations it is evident that a deep understanding of the acoustic channel response at frequencies higher than 2 MHz and for distances longer than few centimeters is still missing and will be the focus of this work.

2.2 Ultrasounds Physics

Similar to light, acoustic vibrations travel in the form of a wave. However, although other waves, like light waves, can travel in vacuum, acoustic waves require an elastic medium such as a liquid or a solid to propagate.

With respect to the entire acoustic spectrum, ultrasounds are mechanical vibrations covering frequencies greater than 20 kHz, far beyond the human hearing range.

In a previous work [26] we have investigated on the suitable range of frequencies to support transmission in tissues. In particular we have identified that, given that attenuation and frequency are inversely proportional, depending on tissues features, there is a maximum allowed carrier frequency for a given maximum tolerable attenuation. In particular, in order to guarantee a suitable communication distance (in the order of tens of centimeters) we observed that the transmission frequency should not exceed 10 MHz. Similar results were also derived independently in [33]. Accordingly, in this paper we will consider the 1-5 MHz spectrum range since it represents a good trade-off between a reasonable attenuation, as discussed above, and a good transducer size, as discussed in the following.

Propagation of acoustic and, specifically, ultrasonic waves, is impacted by a number of features. At a high level, the main problems exhibited during pressure wave propagation can be classified in:

- *attenuation* of the transmitted signals, related to the absorption of sound

waves in media and, specifically, in environments with high water concentration;

- *signal perturbation and deformation* due to variations in the sound speed caused by an heterogeneous medium. Reflections on the material surfaces can also cause strong multi-path, generating parasite echoes and interference;

- *environmental noise inside the body* due to ongoing biological mechanisms inside the body.

In terms of communication issues, this result in absorption, reflection, scattering, refraction, diffraction and interference.

- *Absorption* occurs when part of the sound energy is converted into heat from a medium. This obviously introduces loss;

- *Reflection* occurs when a sound wave is back-propagated because of an interface between two different media. Reflection can be *specular* if the reflected wave is characterized by a single direction, or *diffuse* in case of multiple directions

- *Scattering* causes a redistribution of the acoustic energy among different wave numbers and frequencies resulting in spectral and directional broadening of the waves at high frequencies;

- *Refraction* corresponds to a change of the sound beam directivity induced by the mean velocity gradient of the flow;

- *Diffraction* occurs when the sound wave meets a large obstacle so that a shadow is generated behind, where the incident wave could not penetrate deeply;

- *Interference* happen when different acoustic sources emit sound waves at the same frequency and their distance equals a wavelength multiple. Briefly, interference appears where multiple waves are superimposed.

In the following wave propagation and particle motion will be described.

2.2.1 Wave Propagation and Particle Motion

Acoustic wave propagation consists of a sequence of compressions and dilatations into an elastic medium. The propagation rate of the consequent medium perturbation is denoted as *acoustic wave velocity*. The propagation velocity of an acoustic wave depends on the propagation medium as

$$c = \sqrt{\frac{E}{\rho}} \tag{2.1}$$

where E is the elastic modulus that quantifies the relative variation of volume or density as a function of pressure variations. ρ is the medium density.

At a given temperature and pressure, in a particular point in space, the c velocity in a perfectly elastic material is independent of temperature and pressure , i.e. $c(T, p) = c$, and the relationship between velocity, frequency f and wavelength of a wave λ is:

$$\lambda = \frac{c}{f} \tag{2.2}$$

However, an ultrasonic source generates more than one mechanical wave. In particular, four types of waves can be distinguished (Figure 2.1):

- The *Longitudinal wave,* that is a compression wave in which the particle motion and the wave propagation share the same direction;

- The *Shear wave* where the particle motion and the direction of the propagation are perpendicular;

- The *Surface (Rayleigh) wave,* that travels through the surface of a material and has an elliptical particle motion. Its depth of penetration is approximately equal to one wavelength and the velocity is approximately 90% of the shear wave velocity of the material;

- The *Love wave* that has a complex vibration and is parallel to the plane layer, perpendicular to the wave direction.

All these waves induce pressure variations over the three dimensions and can be fully described by the Helmholtz equation:

$$\nabla^2 P - \frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} = 0 \tag{2.3}$$

where $P(x,y,z,t)$ is the pressure field of a wave propagating in space as a function of time, and $c(x,y,z)$ is the local material propagation velocity that, in general, can be dependent on the position.

If the ultrasonic wave encounters an absorbing medium, the initial pressure P_0 decays following the below relationship:

$$P(d) = P_0 e^{-\beta d} \tag{2.4}$$
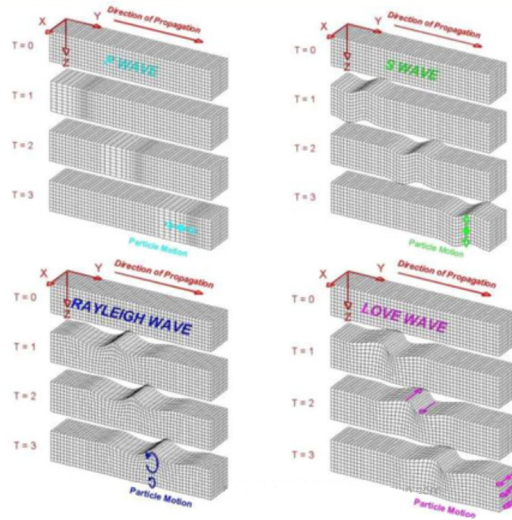
Figure 2.1: a) Longitudinal (particle motion and wave propagation in the same direction), b) Shear (orthogonal particle motion and wave propagation) , c) Rayleigh (elliptical particle motion and wave propagation across the surface) and d) Love (particle motion parallel to the plane layer and orthogonal to wave propagation) waves ([34]).

where d is distance and β is the amplitude attenuation coefficient.

Clearly, in case of non-isotropic environments, multiple set of waves appear due to multipath fading. Their number and entities are related to the particular scattering phenomenon generated in the material.

2.3 The Ultrasonic Transducer

An ultrasonic transducer is a device used to generate and detect ultrasonic waves. Different types of ultrasonic transducers are available, depending on the specific application they are designed for. Contact, immersion, and air-coupled transducers are some of the most common ultrasonic devices employed in many applications. They mainly differ on the coupling media placed between the transducer and the sample. Contact transducers employ

a highly viscous fluid for coupling so that a reduced mismatch between the impedance of the tested material and the transducer is achieved; immersion transducers use instead water for coupling. Air-coupled transducers use air for the waves to propagate.

The use of a transducer implies a good knowledge of its physics. Each of its parameters influences the propagation of acoustic waves inside the material. In Figure 2.2, the main components of a generic ultrasonic transducer are shown. It consists of an active element, a backing and a wear plate. The active element represents the most relevant bit, where a piezo or ferroelectric material trasduces the electrical energy into a mechanical wave.

Once the ultrasonic wave leaves the wear plate, two propagation regions can be identified: the *near field* (or *Fresnel zone*) and the *far field* (or *Fraunhofer zone*). In the near field a series of maxima and minima of the echo amplitude occur until the last maximum is reached, at distance N (see Figure 2.3).

Then, in the far field, the beam profile spreads and the sound field pressure gradually decreases.

The near field distance N depends on the transducer diameter D, the propagation velocity of the acoustic wave c and the frequency f as follows:

$$N = \frac{D^2 f}{4c} \tag{2.5}$$

Of note, according to eq. (2.5) the higher the frequency or the transducer diameter, the larger the near field region.

The far field beam spread can also be characterized by its divergence angle
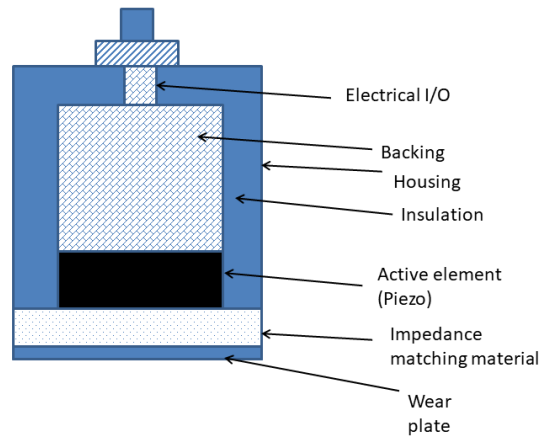
Figure 2.2: Main components of an acoustic transducer.

according to the following equation:

$$\sin \frac{\gamma}{2} = \frac{0.514v}{fD} \tag{2.6}$$

where $\frac{\gamma}{2}$ is the half -6dB angle spread.

This angle measures the divergence of the beam starting from the central axis to the point where the sound pressure has decreased by one half (-6 dB). The shape of the beam and, in particular, the near and far field regions for a particular material, can thus be appropriately tuned as a function of the transducer size and frequency.
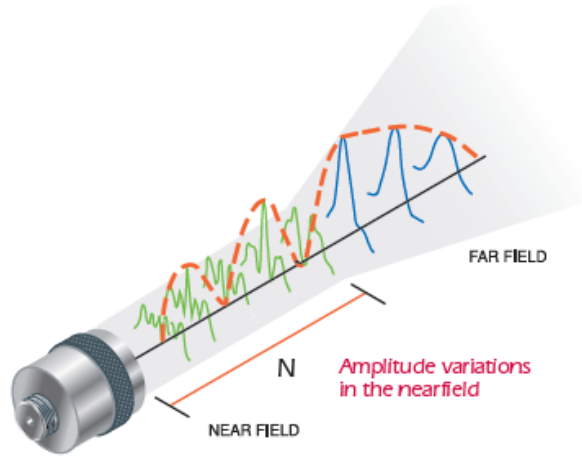
Figure 2.3: Near and Far Field [35].

2.4   Channel Characterization

In order to completely characterise the wireless ultrasonic channel, in this section we will derive the impulse response. As in other wireless channels, this can be mathematically modeled as a linear filter with a complex low pass equivalent impulse response in the form of [26, 36]:

$$h(t) = \sum_l^L \alpha_l \delta(t - \tau_l) \tag{2.7}$$

where α and τ are the complex gain and delay for each of the l replicas and L is the overall number of replicas due to multipath, i.e. $l \in \{1 \dots L\}$. Indeed, in a wave propagating a medium, at least three components can be identified:

- *Direct wave*, which travels through the medium in Line-of-Sight (LoS) from the transmitter to the receiver;

- *Lateral wave*, which propagates out of the medium, along the external surface and, then, enters the medium to reach the receiver;

- *Reflected wave*, which travels through the medium and is reflected at the interface between the medium and other tissues such as bones, muscles or external air.

Accordingly, in eq. (2.7), among the L components we will denote L_D the number of direct wave components, L_R the number of reflected components and L_L the number of lateral components, each characterised by its own gain and delay.

Another important parameter used to characterize a channel is the *path loss*, defined (in logarithmic form) as the difference between the transmitted power (at the end of the transmission system) and the received power (at the end of the receiver antenna, before entering the receiver circuit), i.e.

$$PL = P_t - P_r \tag{2.8}$$

In the above eq. (2.8), the received power[1] is given (in logarithmic form) as the sum of the received power of the three types of waves (direct, reflected and lateral waves) and thus

$$P_r = \sum_{l_D}^{L_D} |\alpha_{l_D}|^2 + \sum_{l_R}^{L_R} |\alpha_{l_R}|^2 + \sum_{l_L}^{L_L} |\alpha_{l_L}|^2 \tag{2.9}$$

In order to derive the channel impulse response, the *excess delay* and the *mean excess delay* are useful metrics that can be derived. The former quantifies

---

[1]We are assuming that $P_t = 1$ W.

the time elapsed between the first and the last arriving components, while the latter is the first moment of the power delay profile and can be derived as:

$$\tau = \frac{\sum_k P_k \tau_k}{\sum_k P_k} \tag{2.10}$$

where τ_k is the delay of the k-th replica and P_k is the associated instantaneous power.

Finally, the *root mean square (RMS) delay spread* is also calculated to give a complete description of the channel. This parameter is considered as a good indicator of multipath spread, because it gives hints on the occurrence of a possible inter-symbol interference (ISI).

The RMS delay spread is defined as:

$$\tau_{RMS} = \sqrt{(\tau^2) - (\tau)^2} \tag{2.11}$$

where (τ^2) is

$$(\tau^2) = \frac{\sum_k P_k \tau_k^2}{\sum_k P_k} \tag{2.12}$$

and τ_k is the associated delay for the k-th replica.

The the above metrics being defined, in the following sections we will detail the experimental procedure used to derive the channel impulse response in a real setting.

17

2.5 Measurement Procedure

The measurement procedure required use of a set of laboratory instruments. The signal input is provided by an Agilent 33220A waveform generator [37] that produces a 10 V peak-to-peak Voltage pulse with 9% duty cycle and null offset. Via BNC connections, this pulse is delivered to a V326-SU Olympus transducer [35] that emits the resulting ultrasonic wave through an artifical human body mimicking phantom. The latter consists of 10% ballistic gelatin, used to mimic human muscle tissues [38]; we also considered a set of experiments where an animal bone is encapsulated in the gel in order to test an even more realistic scenario for BANs [2]. Many studies has been conducted on ultrasounds propagation speed in phantom tissues. Based on experiments described in [39], [40] and [41] it has been observed that the ballistic gel sound velocity is slightly lower than in water, exhibiting values around $1450\,\mathrm{m\,s^{-1}}$. The ultrasonic pulse propagates inside the gel block in such a way to generate multiple copies, as foreseen and discussed above. Then, a receiving transducer located at the opposite side of the ballistic gel block, sends the received signal to a Mini-Circuits ZFL-1000LN+ Low Power Amplifier (LNA) [42] connected to a Keysight Infiniium oscilloscope 900A series oscilloscope [43].

The latter estimates amplitude and delay of the received pulse.

An illustrative example of the laboratory setup is shown in Figure 2.4.

To control the overall system we used the National Instrument LabVIEW software [44] to develop a customized Virtual Instrument (VI) and control

---

[2]This work has respected the European Commission Guidelines on scientific activity involving animals. In particular, organic tissues (e.g. bones) have been obtained from those commercially available.

either the waveform generator and the oscilloscope. In Figure 2.5 the front panel of our VI is shown.

In particular, the VI sets the generation of a pulse train at a defined frequency and queries the oscilloscope about the received amplitude. The frequency varies from 1 MHz to 5 MHz with a 1 kHz step. The choice of this set of frequencies is motivated by our preliminary work in [14, 26] where we showed that this range guarantees an appropriate trade-off between a controllable transducer size, low tissue attenuation and a high directivity. After repeating each set of measurements one hundred times, a table reporting the transmitted signal amplitude, the received signal amplitude and the frequency used for each iteration has been obtained.

We performed the measurement campaign in different settings. In particular, we varied the size and the composition of the medium considering a ballistic gel block with three possible widths, namely 5.5 cm, 10 cm and 12 cm. Then, for each size, we considered not only the pure ballistic gel case, but also the condition where an animal bone 2 cm × 5 cm × 0.5 cm in size is encapsulated in the gel. Our results provide an accuracy of 97.5%. Figure 2.6 illustrates the ballistic gel block where a bone is embedded.

The size of the block was chosen taking into account the transducer properties. Indeed, by remembering eq. (2.5) and given a transducer diameter of 952 mm [35], we calculated the N value for 3.7 MHz which is perfectly compatible with the central frequency declared in the transducer datasheet, by also taking into account that the presence of the LNA partially distorts
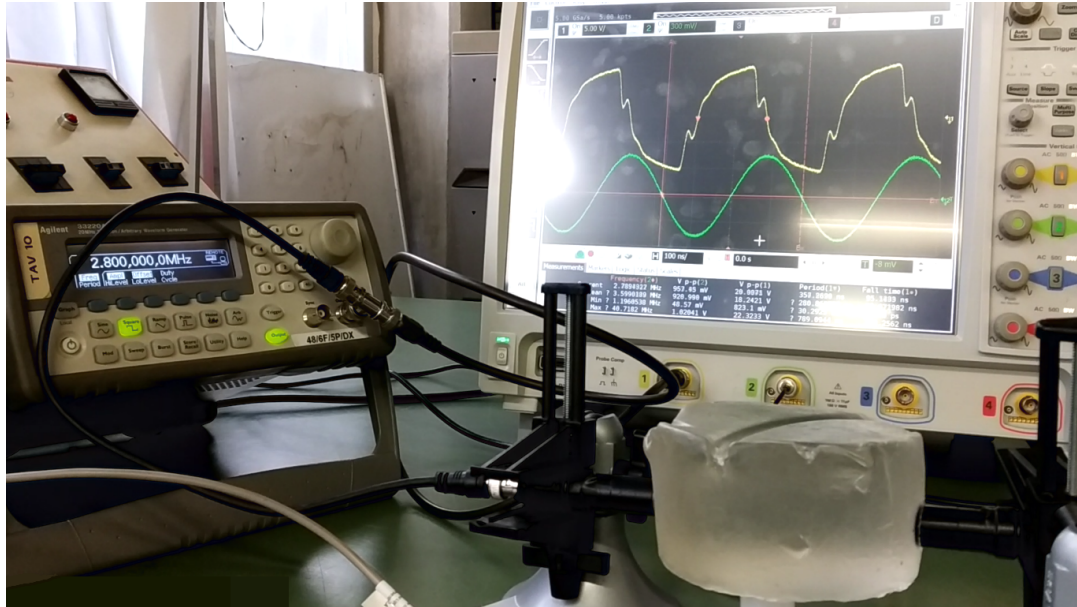
Figure 2.4: Experimental setting: Waveform generator, transmitter and receiver transducers, body phantom made of ballistic gel and oscilloscope.

the signal, thus reducing the central frequency with respect to the declared value. Accordingly, the desired near field distance is approximately 5.7 cm. The other block sizes (10 cm and 12 cm) were considered in order to test the possibility of longer distances.
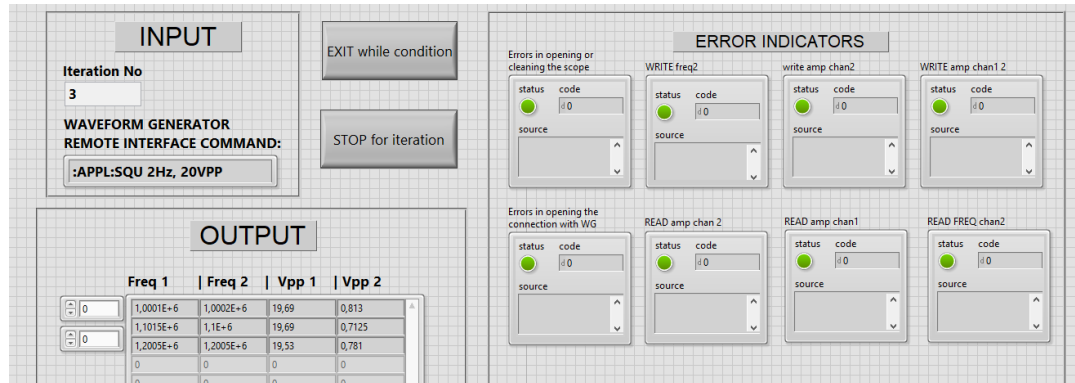
Figure 2.5: Customized Virtual Instrument developed in LabVIEW.

Figure 2.6: Ballistic gel with embedded bone.

2.6 Numerical Analysis

In the following we will discuss the impact of channel composition and distance on different performance metrics such as delay and frequency response. Tables 2.1 and 2.2 summarise all the experimental parameters obtained for each configuration. We will refer to them in the below considerations [3].

[3]In doing our experiments we assumed that a perfect matching of impedance happens and disregarded mismatch or microscopic phenomena that could arise in the derivation of

	5.5 cm				10 cm				12 cm			
	Homo		Hetero		Homo		Hetero		Homo		Hetero	
	α(dB)	Av τ(μs)	α(dB)	Av τ(μs)	α(dB)	Av τ(μs)	α(dB)	Av τ(μs)	α(dB)	Av τ(μs)	α(dB)	Av τ(μs)
1st component	-9.51	37.47	-14.90	29.46	-16.98	70.05	-17.21	68.82	-17.55	80	-25.43	84.36
2nd component	-52.39	112.8	-40.93	89.51	-50.4	192	-47.27	967.92	-41.98	240.76	N.A.	250
3rd component	N.A.	187	N.A.	N.A.	-53.36	315.8	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Table 2.1: Gain module and delay of each component in all configurations

2.6.1 Impulse Response

To derive the impulse response we recall the eq. (2.7), which can be characterized upon specifying the gain (α) and the delay (τ) associated to each replica. At every measurement campaign, we send a pulse with 9% duty cycle and null offset; accordingly, the received signal consists of three multipath replicas, i.e. the direct, surface and reflected waves. When the last component exhibits a power level comparable with noise, the weak replica cannot be detected by the receiving transducer. This occurs when distances are large (e.g. 12 cm) or a heterogeneous channel (ballistic gel with bone) significantly attenuates the power of each component.

Figure 2.7 shows the gain modulus in case of homogeneous channel (i.e. pure ballistic gel) and a 10 cm gel block size. Note that the direct component reaches the transducer after 70.5 μs with a peak amplitude of -16.98 dB; the second component has a delay equal to 192 μs and its peak amplitude is -50.4 dB. Finally, the last component reaches the receiver after 315.8 μs with a -53.36 dB peak amplitude.

Table 2.1 summarizes all α and τ values for the considered configurations.

the impulse response and all parameters discussed in the rest of this section.

	5.5 cm		10 cm		12 cm	
	Homogeneous	Heterogeneous	Homogeneous	Heterogeneous	Homogeneous	Heterogeneous
Path Loss (dB)	9.51	14.90	16.98	17.21	17.55	25.43
Excess Delay (µs)	75.33	60.05	121.95	899.1	160.76	165
# of replicas	3	2	3	2	2	2
τ (µs)	37.195 ± 0.27	29.595 ± 0.02	70.269 ± 0.005	69.624 ± 0.031	80.293 ± 0.03	83.545 ± 0.05
τ_{RMS} (µs)	6.4015 ± 0.0005	30.20 ± 0.04	6.512 ± 0.768	26.79 ± 0.02	6.922 ± 0.289	N.A.

Table 2.2: Main impulse response parameters for all configurations
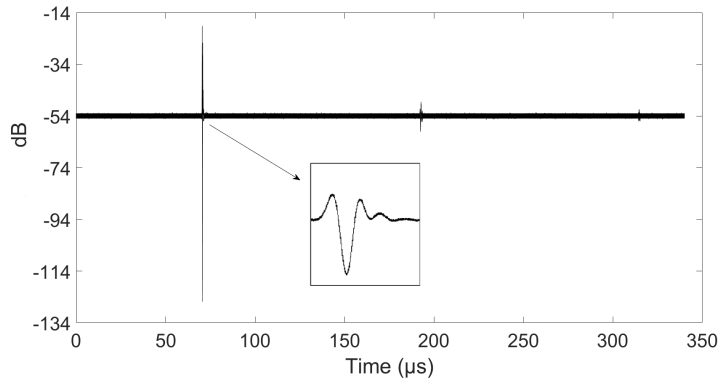
Figure 2.7: Gain modulus in case of homogeneous channel and 10 cm gel block size.

2.6.2 Path Loss

In this section we investigate on the path loss. We calculated the path loss for each considered scenario, i.e. both Homogeneous and Heterogeneous channel, and three different gel block sizes. We observed that, as expected, the path loss increases with distance and with the addition of discontinuities in the medium.

As indicated in Table 2.2, in case of Homogeneous channel, the path loss seems to increase approximately at an average rate of 0.972 dB/cm.

The addition of a discontinuity element (i.e. heterogeneous case) makes the path loss decrease in a non-linear manner, reaching high values at larger

distances. Note that in particular, at 12 cm, the gain modulus is -25.43 dBm while at 10 cm it was -17.21 dBm.

2.6.3 Excess Delay

In this paragraph we investigate on the excess delay which is defined as the time interval between the arrival of the first and the last signal components. Based on this definition, it is clear that the estimation of this parameter is affected by the number of copies obtained at the receiver. Accordingly, in order to take into account the lack of received components in case of larger distances, we considered only the time interval between the reception of the first and the second components. In case of Homogeneous channel, at 5.5 cm gel block size the excess delay is 75.33 µs, at 10 cm it is almost doubled, i.e. 121.95 µs, while for 12 cm the value of the excess delay is 160.7 µs. These values suggest that the excess delay increases approximately at a rate of 10.26 µs/cm.

Observe also that, as expected, the larger is the block size — and thus the distance between transmitter and receiver — the higher is the difference between the time employed by the direct wave to reach the destination and the time needed by the reflected components.

In case of Heterogeneous channel, the presence of scattering effects seems to make this parameter unpredictable with respect to distance. Indeed, the measured values of the excess delay are: 60.0 µs for 5.5 cm gel block size, 899.1 µs for 10 cm and 165 µs for 12 cm. Note that, in case of 10 cm, due to the unpredictable effects of a not perfectly identical positioning of the animal bone with respect to the other two cases, the scattered components are reflected

multiple times, thus causing an excessive and unpredictable increase in the delay.

2.6.4 Mean Excess Delay

We have observed so far how the delay spread metrics are influenced by the number of received copies. Similar considerations can be done for the mean excess delay and for the RMS delay spread reported in Table 2.2.

Table 2.1 shows how τ depends on the number of replicas and the gel block size. Note that, the presence of discontinuity elements in the communication channel (i.e. heterogeneous case) does not necessarily correspond to an increased value of the mean excess delay. In fact, multipath copies could sum to superficial waves in a positive or destructive manner.

In Figure 2.8 we show the Cumulative Distribution Function of τ in case of 10 cm gel block size both for Homogeneous and for Heterogeneous channel. Note that the positive summing effect due to the presence of the discontinuity element at 10 cm is evident by observing the set of the mean excess delay values. However note that the delay associated to the direct wave (i.e. the first component) is approximately unchanged.

2.6.5 RMS Delay Spread

In this paragraph experimental measurements on RMS delay spread are provided. From Table 2.2 observe that τ_{RMS} seems to slightly depend on the distance between transmitter and receiver, while it is mainly impacted by the channel composition.
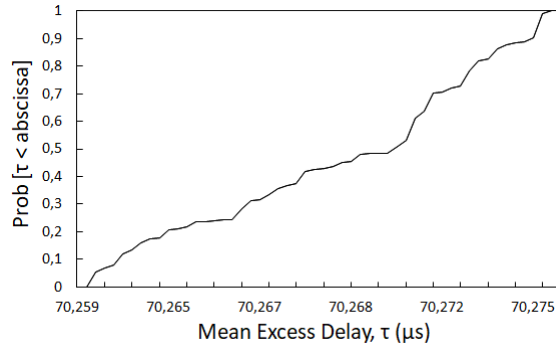
(a)
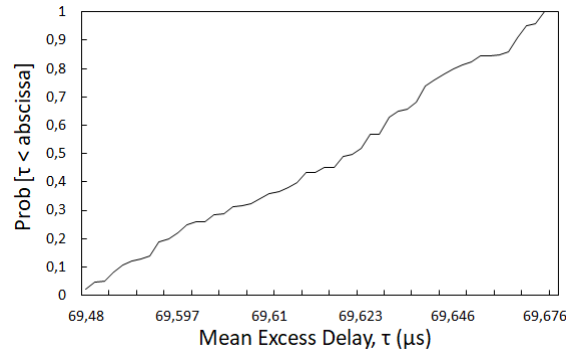
(b)

Figure 2.8: Distribution of Mean Excess Delay τ at 10 cm gel block size in both homogeneous (a) and heterogeneous (b) channels.

Both for 5.5 cm and 10 cm block sizes, observe that the RMS delay in the homogeneous case is about one fifth as compared to the heterogeneous case. This means that the effect of having a homogeneous channel is to reduce the variability in the delay.

Unfortunately in case of 12 cm gel block size, the RMS for the heterogeneous case cannot be recorded because in this case only one component arrives at the receiver. As previously stated, in case of higher distances and heterogeneous channel, each component is heavily attenuated; so in this particular case, even the second component has a low amplitude, comparable with noise
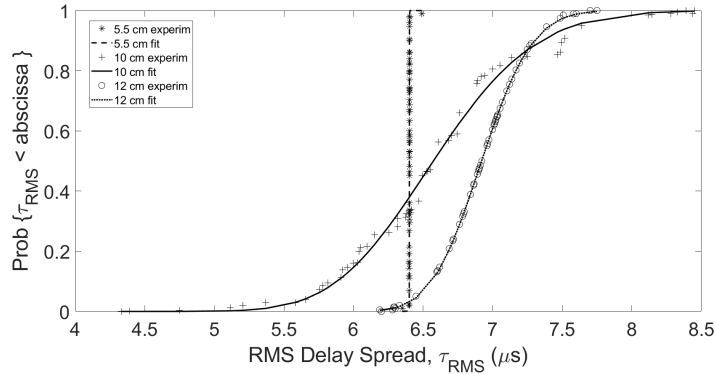
Figure 2.9: CDF of the RMS delay spread in case of homogeneous channel for 5.5 cm, 10 cm and 12 cm gel blocks.

and, thus, cannot be recorded.

In Figure 2.9 we report the Cumulative Distribution Function of τ_{RMS} in case of homogeneous channel for three distinct gel block sizes. It is clear that at lowest distance this parameter is more stable showing a standard deviation σ equal to 0.0005 µs. This is because when the transmitter-receiver distance is low, in spite of the different path traveled by the three components, the delay remains limited. In the other cases instead, upon increasing the distance between the transmitter and the receiver, the difference in the traveled path becomes relevant and the RMS delay spread increases. In the extreme case, when only two components are received (i.e. gel block size 12 cm), the RMS delay spread evaluation is imprecise and suffers for the lack of part of the delay contribution.

Note that in the same figure we also illustrate the comparison between the experimental data and the log-normal distribution. As evident, the RMS Delay Spread can be described quite well by the log-normal fit.

For the homogeneous channel described above, it is possible to calculate the coherence bandwidth of the channel which is proportional to the inverse of the RMS delay spread τ_{RMS}. Accordingly, we could consider this channel as frequency- selective for signals of bandwidth above approximately 30 kHz [45].

2.6.6 Effect of Frequency

In this paragraph we perform an analysis of the gain modulus (i.e. $|\alpha|$) as a function of the frequency used.

In particular, we plot the transfer function modulus (i.e. the attenuation) depending on the frequency being selected in the experiments which has been chosen in the range 1.6-5MHz based on the considerations done in previous sections.

In Figures 2.10 and 2.12 we show $|\alpha|$ for each of the three gel block sizes considered, upon issuing either a square wave or a pulse. The square wave has 10 peak-to-peak Volt amplitude, 50% duty cycle and null offset. The pulse has 10 peak-to-peak Volt amplitude, 9% duty cycle, null offset. In the same plot we report also the standard deviation and the confidence interval (in grey color) set to 97.5%. In Figures 2.11 and 2.13 the same analysis is performed in case of heterogeneous channel.

By comparing Figures 2.10 and 2.11, as expected, we observe that the use of heterogeneous channels with a higher density causes an increase in $|\alpha|$. This is also evident by considering Figures 2.12 and 2.13. Note that, in case of
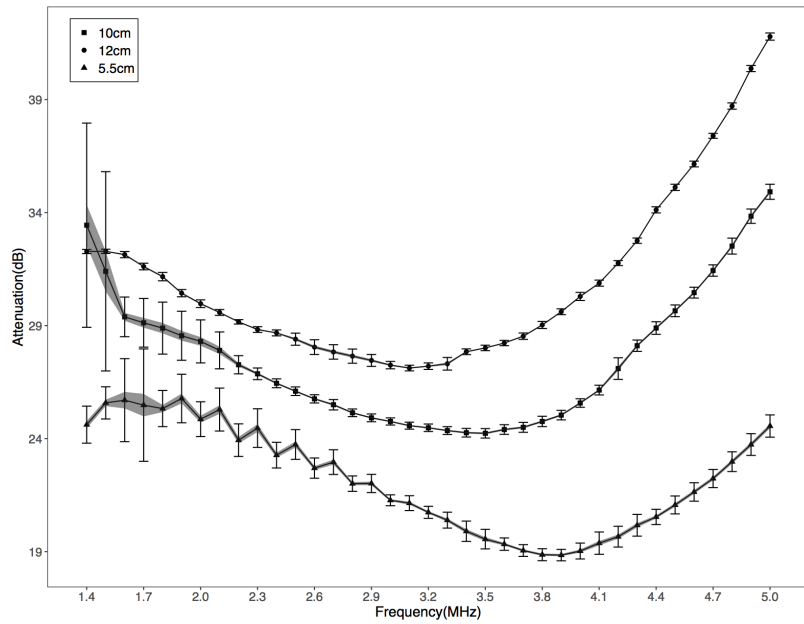
Figure 2.10: $|\alpha|$ Vs. Frequency in homogeneous channel when issuing a square pulse.
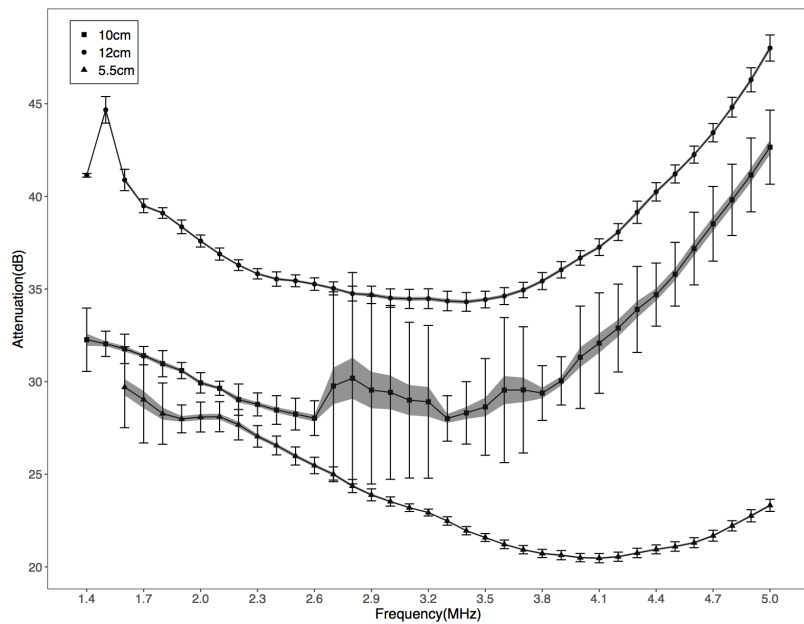
Figure 2.11: $|\alpha|$ Vs. Frequency in heterogeneous channel when issuing a square pulse.
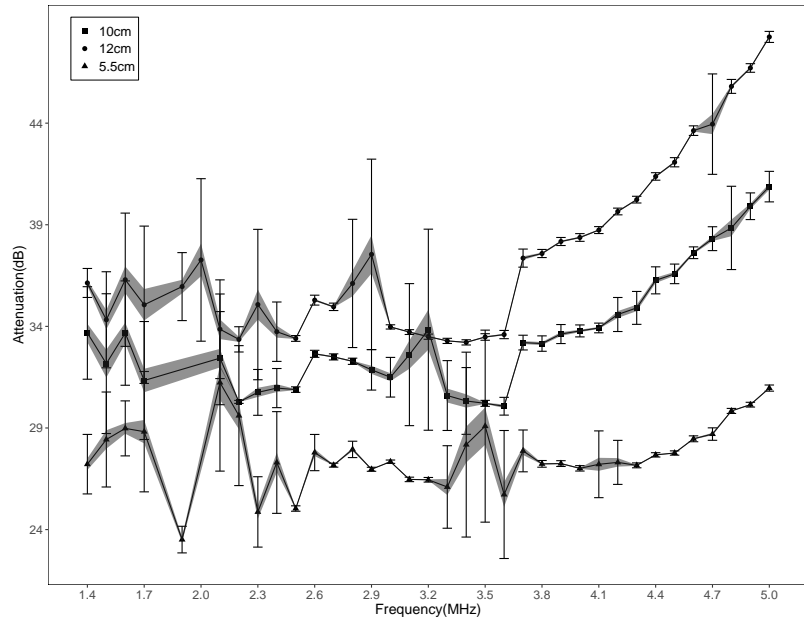
Figure 2.12: $|\alpha|$ Vs. Frequency in homogeneous channel when issuing a pulse.
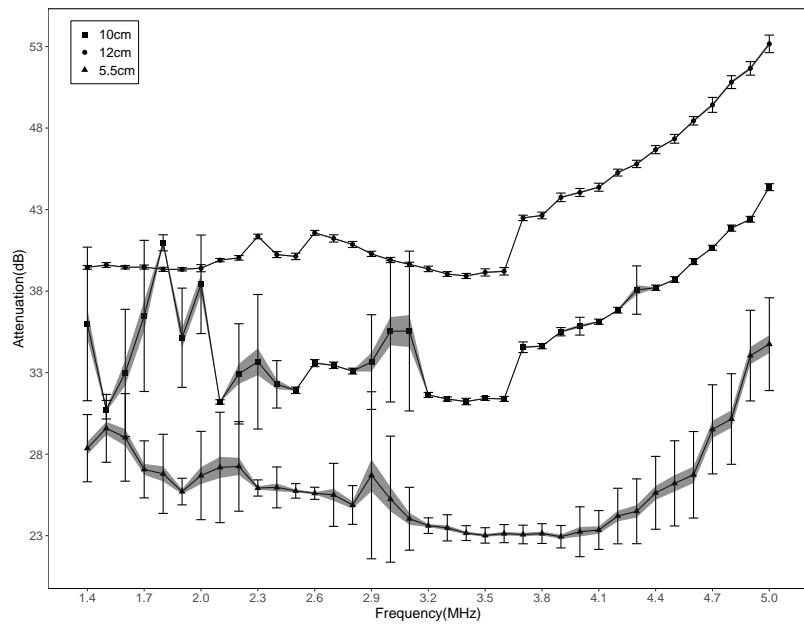
Figure 2.13: $|\alpha|$ Vs. Frequency in heterogeneous channel when issuing a pulse.

a heterogeneous channel, the standard deviation increases significantly too, especially in case of 10 cm which has shown to be the most critical size for the gel block size. Also note that, due to the features of the ultrasonic transducer being employed, i.e. the V326-SU Olympus transducer, which has a maximum sensitivity at 3.4-3.5 MHz, $|\alpha|$ is minimal at these frequencies. We also observe that, due to the small coherence bandwidth, in case of pulse wave, the high signal frequency leads to larger variations in the attenuation.

2.7   Effect of propagation inside tissues

In this section we investigate on the dynamics of ultrasonic signal propagation inside tissues. Specifically we developed k-Wave simulations [46] of our testbed system consisting of a cylinder of ballistic gel with a bone embedded in it. In Figure 2.14 we show the considered topology, the signal received at different positions inside the bone, namely at the ingress of the bone, in the intermediate positions at 5 mm of distance each, till the egress of the bone. In Figures 2.15 (a) and (b) we observe the intensity of the received signal at the different positions inside the bone and at different time instants. Note that the maximum received signal intensity decreases upon moving rightmost inside the bone. This is because, as expected, the bone tissues add an attenuation to the signal. In Figure 2.15b a graphical representation of the intensity is shown with colors ranging from red to grey. Note the drastic decrease in the received signal at the different interfaces, in particular at the egress point of the bone.

In Figures 2.16 three snapshots report the ultrasonic acoustic field at different time instants and a different positions, inside the phantom and in

31
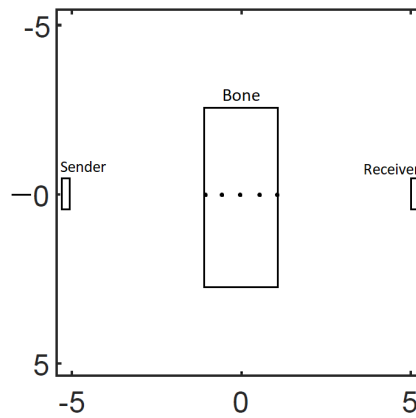
Figure 2.14: 2D topology considered for k-Wave simulations.

proximity of the bone. Note that the effect of the presence of the bone is to cause numerous reflected waves and, thus, the presence of bone and similar heterogeneous tissues as compared to homogeneous ballistic gel, can turn into an advantage because it allows, through the existence of reflections, to propagate the signal also to far-away destinations.
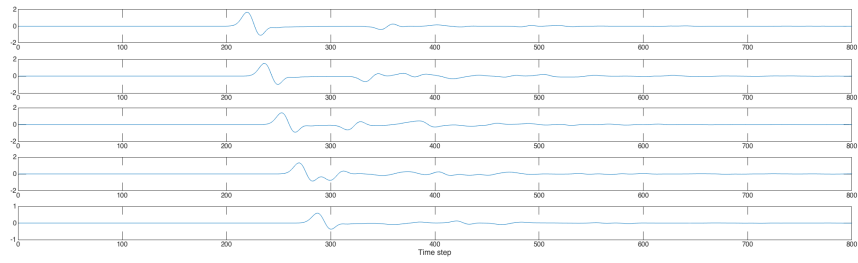
(a)
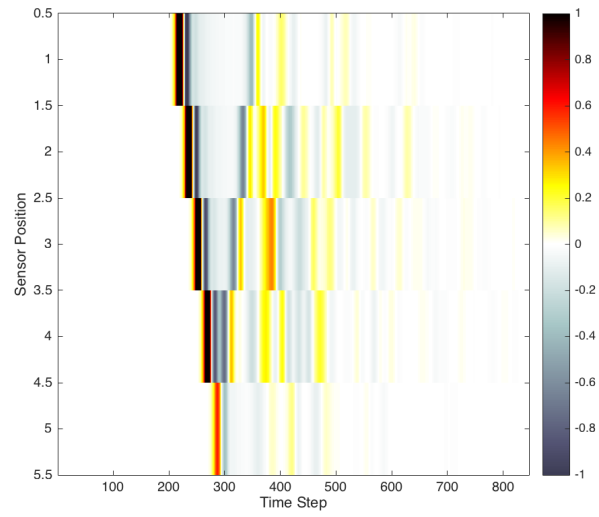
(b)

Figure 2.15: Received Ultrasonic Signal Intensity at different positions across the bone tissue as a function of time (a) and as a function of position and time (b).
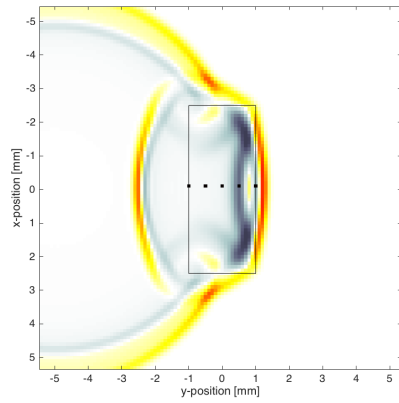
(a)
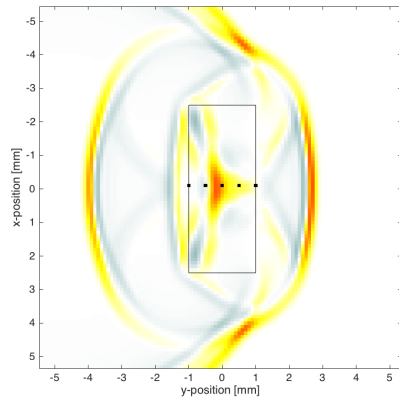
(b)
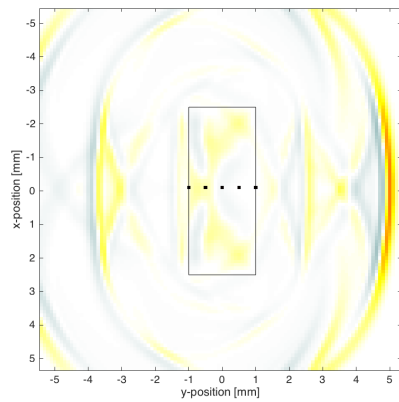
(c)

Figure 2.16: Ultrasonic field as a function of the distance traveled at three different time instants.

2.8 Chapter Summary

In this chapter, the use of ultrasounds has been introduced for communications across intra-body sensors.

To allow a reliable communication, an impulse response analysis of a channel comparable to the human body has been presented and assessed using experimental data. Distribution of mean excess delay and RMS delay spread has been investigated showing that they can be fitted by a log-normal distribution. We have discussed the effect of considering an homogeneous channel or a heterogeneous one where components with different densities have been added. Also the effect of transmitter-receiver distance has been investigated. The coherence bandwidth of the channel has been estimated in relationship with the RMS delay spread. The behavior of the channel at different frequency ranges has been studied as well as the propagation of the ultrasonic wave across heterogeneous tissues. Results provided in this work can be considered as a preliminary useful tool for the design of intra-body communication systems for precision medicine.

Chapter 3

New methods for RNA-sequencing data analysis

The content of this Chapter is partially included in the abstract published on Rheumatology as [47].

While treatment fine-tuning can be achieved by constant monitoring of selected parameters, preliminary studies on prediction of response should be set for those pathologies where multiple therapies are available.

Even for long known diseases as for example rheumatoid arthritis, treatment options have improved substantially with both synthetic and biologic therapies. Yet, treatment response remains low with around 40% of patients failing treatments.

The identification of gene expression levels patterns has proven to be a promising way to stratify patients and therefore help clinicians in the decision making process of a therapy. On the other hand, Next Generation Sequencing (NGS) technology opened the doors to a vast quantity of RNA-sequencing data increasing chances of getting genetic signatures associated with phenotypes.

In in the first section of this Chapter (3.1) we will introduce a new network-based tool to analyse RNA-seq data and highlight the activation of biological interactions characterising groups of patients.

In the second section (3.2) a new way of aggregating and visualising pathway alterations across multiple patient groups will be discussed.

3.1 Pat-BINE: A new network-based tool to characterise sub-groups of patients

Many statistical tools have been developed over time to determine characteristic genes related to specific groups of patients. However, a list of genes is not sufficient to understand underlying biological mechanisms. A biologically relevant view is fundamental to analyse how the identified genes are interacting within patient sub-groups. This can facilitate repurposing of drugs and stratified treatment based on interfering somewhere within a pathway rather than only relying on very few genes to target. On the other hand, several tools outputting a list of up or down-regulated pathways are available (e.g. IPA [48], MITHrIL [49], Reactome [50], SPIA [51], PARADIGM [52], etc.). However, they point to predefined pathways without specifying the links that are effectively active into it. In this section we will introduce a novel pathway-based tool called Pat-BINE (Patients Biological Interactions Network Extractor). Pat-BINE exploits the information of four well curated databases to extract sub-networks of relevant biological interactions characterising patient sub-groups.

This allows clinicians not only to have an indication of the altered genes in a

group, but also to explore how they are linked together in a comprehensive network of interactions.

The main difference with a Differentially Expressed Gene (DEG) analysis lies in the network view. While a DEG analysis points to single genes, the described approach looks at differentially active interactions occurring among the highly expressed genes. For this reason, the tool can be used in combination with a DEG analysis allowing to identify those genes originating an up/down-regulation (upstream genes) or genes that will be affected from it (downstream genes).

In the following sub-sections we will describe the developed pipeline in four steps.

Step 1 - Aggregation

Throughout this phase, patient samples are stratified according to a parameter of interest.

In Chapter 4 we will describe two cases of study originated by two respective parameters. In the first one, the patients are divided according to three histopathological classes. In the second case, a treatment response parameter is used to split the cohort into *good*, *moderate* and *none* patients.

As standard NGS RNA-seq data, patients' samples are organised in a count matrix of gene counts per sample. After Variance Stabilizing Transformation (VST) normalisation, the count matrix is separated into as many sub-matrices as groups of samples the stratification defined (three matrices for both studies in this case).

Each sub-matrix is finally reduced to a vector by replacing each row with its

mean value. As a result, the mean gene expression level of each gene for each category is obtained.

Step 2 - Categorical Networks Creation

At this stage, a list of genes with their related average value is available for each category.

These genes can be linked to one another by utilising available pathway information. This implementation uses the KEGG [53] pathway repository enriched with miRNA interactions from miRTarBase [54] and miRecords [55]. Interactions between transcription factors (TFs) and microRNAs (miRNA) are added from TransmiR [56]. If the count matrix does not contain any TFs or miRNAs those interactions are ignored. This yields to multiple networks of biological interactions built on the gene lists found in Step 1. Here, nodes represent biological entities such as genes, proteins or miRNAs, while edges serve as activation or inhibition interactions among them. A weight is inferred to each node according to its average value estimated in Step 1. Therefore, each network shares the same topology (same nodes and edges) and differs on weights attributed to nodes.

Step 3 - Expression-based cut-off and percolation analysis

To reduce the number of nodes and highlight the most relevant interactions, low expressed genes are removed from each category network. This is done by making use of a user-provided threshold. To choose this threshold, a percolation analysis is performed. Commonly used in statistical physics and mathematics, percolation describes the behaviour of network properties

when nodes or edges are progressively removed [57]. Network parameters are usually plotted as a function of the threshold, and the optimal cut-off value is chosen in correspondence of a change of the function slope or with an inflection point. In the presented method we choose the threshold that would have led to the maximum average number of statistically significant interactions across the three networks (i.e. 75th quantile cut-off). Once the cut-off is set, all the nodes whose mean expression is below the threshold are removed from every network. This naturally leads to different sets of nodes being discarded from each network, therefore, at the end of this phase, every network has a different topology.

Step 4 - Remove overlapping links across networks

In the last phase, the category networks are compared, and all the overlapping links are removed. In other words, if the same edge is present in two or more networks, it will be erased from both of them. This final step allows to bring out category-specific interactions avoiding redundant information between two or more categories.

This tool has been applied to the Pathobiology of Early Arthritis Cohort (PEAC) cohort [5]. Details and results will be discussed in Chapter 4.

3.2 A new feature to visualise altered pathways over multiple groups

The tool described in the previous section can give a detailed description of the biological interactions uniquely activated in a selected category of patients.

This comes extremely useful while studying the etiology of a pathology and trying to understand upstream causes and downstream consequences of a gene level alteration.

Anyway, when a more general view is required, a traditional pathway analysis can be satisfying enough to compare distinct category of patients.

Nevertheless, when it comes to compare more than two categories, it is often hard to visualise the different outcomes at a glance. All the traditional pathway analysis tools offer the possibility of comparing only two groups of patients at a time (usually treated Vs control patients).

To cope with this issue we took advantage of the MITHrIL software [49] to develop a tool able to show the alteration of a pathway over three or more groups of patients.

The traditional way of using this software requires the calculation of the Log-Fold-Change of differentially expressed genes between two groups of patients. This can be easily obtained by means of widely spread software such as limma [58] or DESeq2 [59].

Anyway, MITHrIL also allows to compare the gene expression profile of a single patient to its internal null model. This outputs an estimated alteration of every available pathway compared to the random model.

We used this feature to show the different levels of pathways alteration across multiple group of patients.

The details of this approach can be summarised in the following steps:

- *Step 1* - The gene expression profile of each patient is filtered and log2 normalised by means of limma/voom [58, 60].

- *Step 2* - For each pathway, the estimated alteration from each patient is collected.

- *Step 3* - For every pathway, a boxplot is drawn grouping the patient according to a user defined categorisation. This allow to compare the alteration of a pathway of interest over multiple groups at a glance.

This tool has been applied to the Pathobiology of Early Arthritis Cohort (PEAC) cohort [5]. Details and results will be discussed in Chapter 4.

Chapter 4

Clinical Trials

The content of this Chapter (Section 4.1) has been presented at the American College of Rheumatology Convergence 2020 [61]. An extended report will be submitted to Annals of the Rheumatic Diseases (ARD) as [62]. The content of Section 4.2 will be submitted to Science translational medicine as [63].

Find genetic signatures of treatment response is certainly a gold objective for precision medicine. Although laboratory and animal tests can give an important contribution, the relevance of clinical trials cannot be disregarded. They can highlight what does not work in humans and reveal distinct sub-groups of patients.

For this reason, in this chapter focuses on the analyses and results of two clinical trials.

The first Section (4.1) relies on the Pathobiology of Early Arthritis Cohort (PEAC) [5] whose RNA-seq data is available under request at [64]. A complete transcriptomic analysis on this trial can be found at [5] and [65]. We exploited this data to validate the tools described in Chapter 3 and find new insights on

rheumatology arthitis.

In Section 4.2 a transcriptomic analysis of RNA-seq data collected in the TRial of Anti-B-Cell Therapy In patients with primary Sjögren's Syndrome (TRAC-TISS) [6] will be adressed. It represents the first longitudinal transcriptomic analysis on salivary gland biopsies collected over a year time in Sjögren's Syndrome patients treated both with placebo and immuno-modulatory therapeuthic agent.

4.1 Pathobiology of Early Arthritis Cohort

Study of histological pathotypes and clinical treatment response classes by means of network and pathway analysis

Rhumatoid Arthritis (RA) is a chronic inflammatory autoimmune disease that can cause joint pain, damage and disability. [66]

Several studies have tried to characterize the pathology based on gene expression in the disease tissue, i.e. the synovial membrane. However, most of these studies have investigated the gene expression in synovial tissues obtained mostly from joint replacement, thus from patients with well-established or terminal disease, with the additional bias of concomitant medications. Recently, in [5] and [67] a large cohort of patients with early treatment-naïve RA was profiled, describing different histological patterns, i.e. pathotypes (lympho-myeloid, diffuse-myeloid, pauci-immune fibroid) with corresponding molecular signatures including RNA-sequencing of the synovial tissue. Additionally, they showed how response to conventional synthetic Disease-Modifying Anti-Rheumatic Drugs (csDMARDs) correlates to gene modules. In the study described in this chapter, we utilise gene expression data derived

from the same cohort of treatment-naïve early RA patients, but apply the novel pathway-based tool Pat-BINE defined in Chapter 3. This allows to explore a comprehensive network of interactions that are occurring in patients classified according to synovial inflammation (pathotypes) and response to treatment with csDMARDs.

4.1.1   Patient cohort

In total, RNA-seq data of 94 patients with early, treatment-naïve RA fulfilling the 2010 ACR/EULAR Criteria were recruited into the Pathobiology of Early Arthritis Cohort (PEAC) [5, 67]. We removed 11 samples due to poor RNA quality or ungraded histopathology, the remaining 83 were included in our study (see Table 4.1). As previously described1,2, patients have been stratified into three distinct synovial histopathological groups:

1. *lympho-myeloid* (n=49) characterized by lymphoid cell infiltration such as T cells, B cells and plasma cells together with myeloid cells;

2. *diffuse-myeloid* (n=18) with a high prevalence of cells from the myeloid lineage with low numbers in B cells and plasma-cells;

3. *pauci-immune* (n=16) identified by fibroblast expansion while immune cells are sparse.

After baseline treatment naÃŕve biopsy, patients underwent six months of csDMARD therapy and their responsiveness was assessed according to DAS28 EULAR criteria. This resulted in 26 good, 32 moderate and 9 non responders.

	Lymphoid (N=49)	Myeloid (N=18)	Fibroid (N=16)	Total (N=83)	p value
Age (years)	52.3 (16.2)	50.4 (16.5)	53.2 (15.0)	52.1 (15.9)	0.865
Gender					0.608
- F	37 (75.5%)	12 (66.7%)	13 (81.2%)	62 (74.7%)	
- M	12 (24.5%)	6 (33.3%)	3 (18.8%)	21 (25.3%)	
Disease duration (months)	5.9 (3.3)	4.8 (2.6)	7.0 (3.5)	5.8 (3.3)	0.152
RF					0.203
- positive	32 (69.6%)	9 (52.9%)	7 (46.7%)	48 (61.5%)	
- negative	14 (30.4%)	8 (47.1%)	8 (53.3%)	30 (38.5%)	
CCP					0.025
- positive	39 (84.8%)	10 (55.6%)	9 (60.0%)	58 (73.4%)	
- negative	7 (15.2%)	8 (44.4%)	6 (40.0%)	21 (26.6%)	
DAS28	6.2 (1.2)	5.6 (1.2)	5.2 (1.6)	5.8 (1.3)	0.029
ESR (mm/hr)	50.9 (28.1)	37.6 (25.4)	30.8 (27.7)	44.1 (28.4)	0.025
CRP (μ g/mL)	25.0 (26.5)	17.5 (26.0)	14.4 (41.7)	21.3 (29.9)	0.395
TJC	12.6 (7.2)	9.9 (6.3)	10.9 (8.9)	11.7 (7.4)	0.396
SJC	8.4 (5.7)	7.1 (4.3)	5.2 (5.0)	7.5 (5.4)	0.116
VAS	67.9 (24.0)	61.1 (21.7)	57.8 (27.1)	64.5 (24.2)	0.287
HAQ	1.6 (0.8)	1.4 (0.6)	1.6 (0.8)	1.5 (0.7)	0.499

Table 4.1: Baseline demographics of treatment-naïve RA patients recruited into the Pathobiology of Early Arthritis Cohort (PEAC).

4.1.2 GO enrichment analysis

The presented methodology was applied to five categories of patients determined by their synovial histopathology and their response to six months of csDMARD treatment. The categories are:

- *Lympho-myeloids*

- *Diffuse-myeloids*

- *Pauci-immune fibroids*

- *Good responders*

- *Non responders*

Each resulting network is made of a number of easily identifiable clusters. To perform a systematic interpretation of these clusters we used the clueGO Cytoscape plugin [68].

ClueGO exploits several Gene Ontology (GO) and pathway repositories to find the GO terms/pathways associated to a list of genes given in input. Its features are REST enabled, so we analysed each cluster in an automated way via R software.

We summarised the ClueGO results in Table 4.2.

First, the GO terms/pathways associated to each cluster were filtered by BH adjusted p-value < 0.05. Then the table was built on the following columns:

- *Network*, indicating of one of the five networks associated to a group

- *Cluster*, identifying the cluster inside a network (cluster IDs were assigned for each cluster in a network)

- *Associated GO Term*, the GO term/pathway containing the largest number of genes making the cluster

- *Nr. of Associated Genes*, indicating the actual number of the genes contained in the Associated GO Term, over the total number of genes in the cluster

- *% of Associated Genes*, the percentage of genes in the cluster contained in the Associated GO Term

In the following sections findings from each group will be discussed.

Network	Cluster	Associated GO Term	Nr. of Associated Genes	% of Associated Genes
	A	Chemokine signaling pathway	29/31	93.5
	B	Antigen processing and presentation	6/6	100.0
	C	Focal adhesion	31/33	93.9
	D	TNF signaling pathway	20/41	48.8
	E	RIG-I-like receptor signaling pathway	7/8	87.5
Lympho-myeloid	F	Apoptosis	33/65	50.8
	G	MAPK signaling pathway	44/71	62.0
	H	Wnt signaling pathway	13/16	81.2
	I	positive regulation of interleukin-8 production	6/12	50.0
	L	Interleukin-2 family signaling	3/3	100.0
	M	disulfide oxidoreductase activity	4/4	100.0
	A	Focal adhesion	34/45	75.6
	B	PPAR signaling pathway	22/28	78.6
Diffuse-myeloid	C	Dopaminergic synapse	23/30	76.7
	D	EPHA-mediated growth cone collapse	3/3	100.0
	E	Cam-PDE 1 activation	4/4	100.0
	F	Adherens junction	5/6	83.3
	A	Regulation of RUNX1 Expression and Activity	3/3	100.0
	B	Notch signaling pathway	5/5	100.0
	C	Wnt signaling pathway	18/19	94.7
	D	Focal adhesion	42/45	93.3
	E	Ras signaling pathway	49/64	76.6
Pauci-immune Fibroid	F	JAK-STAT signaling pathway	18/18	100.0
	G	TGF-beta signaling pathway	9/14	64.3
	H	AKT-mediated inactivation of FOXO1A	3/3	100.0
	I	Energy dependent regulation of mTOR by LKB1-AMPK	6/6	100.0
	N	Adrenergic signaling in cardiomyocytes	9/11	81.8
	O	Long-term depression	12/19	63.2
	A	PI3K-Akt signaling pathway	34/64	53.1
	B	Signaling by high-kinase activity BRAF mutants	5/7	71.4
	C	ECM-receptor interaction	7/7	100.0
Good Responders	D	disulfide oxidoreductase activity	4/4	100.0
	E	Chemokine receptors bind chemokines	6/6	100.0
	F	Triglyceride catabolism	3/3	100.0
	G	Cam-PDE 1 activation	4/4	100.0
	A	Olfactory transduction	32/50	64.0
	B	Chemokine signaling pathway	25/25	100.0
	C	Cytosolic DNA-sensing pathway	5/6	83.3
Non Responders	D	Human cytomegalovirus infection	24/39	61.5
	E	Wnt signaling pathway	10/15	66.7
	F	Antigen processing and presentation	6/6	100.0
	G	Adrenergic signaling in cardiomyocytes	11/18	61.1
	H	Wnt signaling pathway	13/16	81.2

Table 4.2: GO/Pathway enrichment analysis on network clusters resulting from the applied method.

4.1.3 Results

4.1.3.1 The lympho-myeloid pathotype

As described in the methods section, the lympho-myeloid network is obtained selecting the most active gene-gene interactions of this class. Among them, those links equivalently active in the diffuse-myeloid and pauci-immune fibroid classes were discarded. This allows inspection of specific functions uniquely characterising the lympho-myeloid phenotype.

The network contains 334 nodes and 469 edges. 373 are activator links, while 96 are inhibitor links. 11 gene clusters were identified, and their functions are reflected in the Associated GO Term in Table 4.2.

The six most prominent clusters regulate the chemokine signaling pathway (93.5% of cluster genes, Figure 4.1.A), Antigen processing and presentation (100%, Fig. 4.1.B), Focal adhesion (93.9%, Fig. 4.1.C) TNF signaling pathway (48.8%, Fig. 4.1.D), Tumour Necrosis Factor (TNF) receptor (TNFR) associated factor Ribonuclease A F3-dependent interferon regulatory factor family (IRF) activation pathway (87.5%, Fig. 4.1.E) and Apoptosis (50.8%, Fig. 4.1.F). Functions of the remaining clusters are MAPK (62.0%, cluster G), Wnt (81.2%, Cluster H) signaling, positive regulation of interleukin-8 production (50%, Cluster I), Interleukin-2 family signaling (100%, Cluster L) and disulfide oxidoreductase activity (100%, Cluster M) (see Table 4.2).
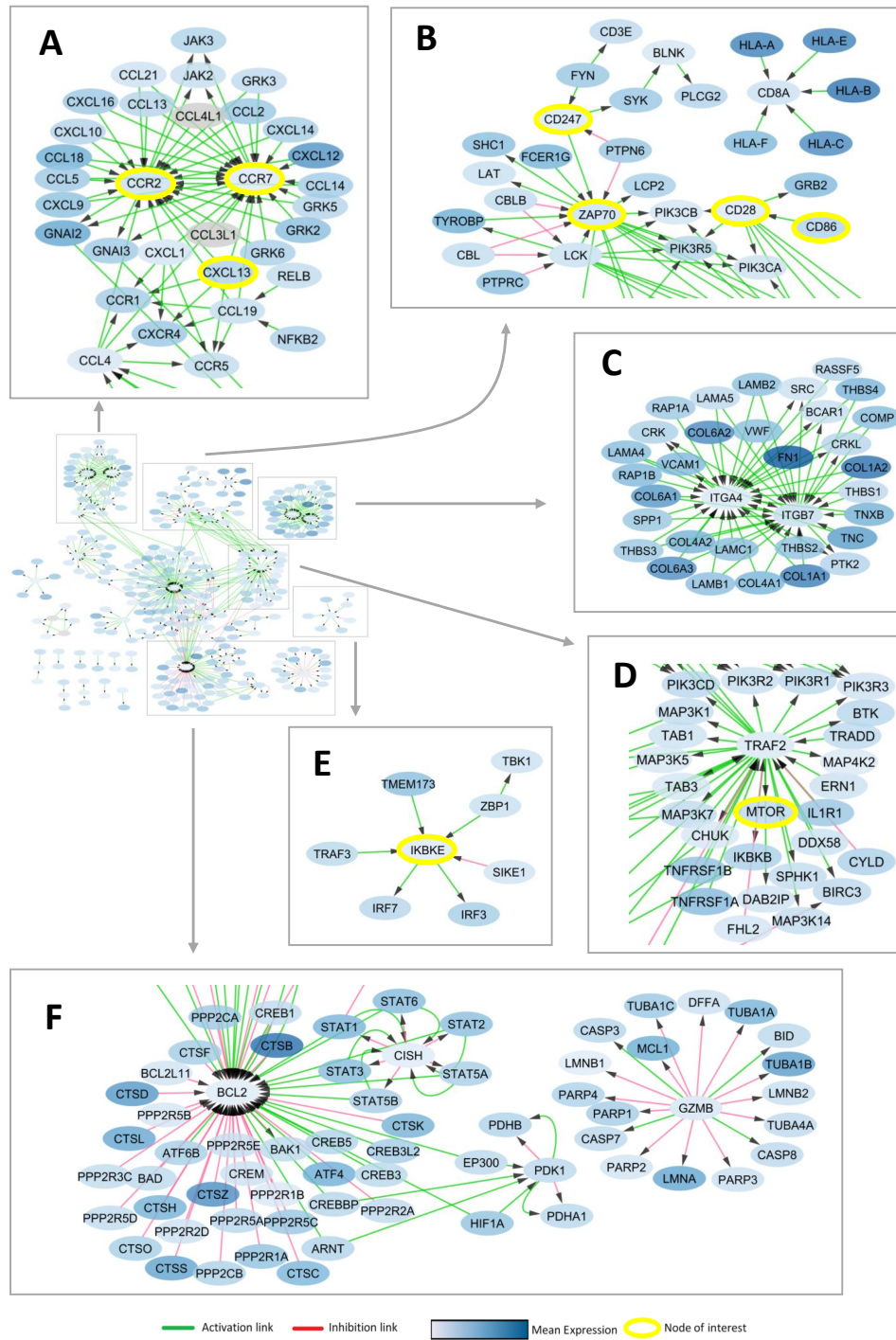
Figure 4.1: (Caption on next page.)

Figure 4.1: (Previous page.) **Unique clusters of genes in joints within the lympho-myeloid pathotype. (A)** Chemokine activation for leukocyte recruitment. **(B)** Antigen processing and presentation with T cell activation genes. **(C)** Group of Focal adhesion genes comprising collagens, integrin and laminin. **(D)** TNF signaling through mTOR. **(E)** Interferon regulating signaling. **(F)** Genes of the intrinsic and extrinsic apoptotic pathways.

4.1.3.2 The diffuse-myeloid pathotype

Among all the active interactions found in the diffuse-myeloid pathotype, those in common with the lympho-myeloid and the pauci-immune fibroid pathotype were removed. This results in only 129 unique nodes with fewer connections compared to the previous pathotype. 98 edges are active connectors while 22 are inhibitors.

The smaller size of this network may be explained by the overlapping characteristics of both the lympho-myeloid and the pauci-immune fibroid group. On the one hand this category is characterized by the infiltration of macrophages, like observed in the lympho-myeloid pathotype; on the other hand, the absence of B and plasma cell aggregates is in common with the pauci-immune fibroid group.

One cluster with the same Associated GO Term as observed in the lymphoid group was for focal adhesion (75.6% Fig. 4.2.A(I)). However, the cluster in the diffuse-myeloid pathotype was centered around ITGB3. On the other hand, unique for the diffuse-myeloid pathotype were Peroxisome proliferator-activated receptor (PPAR) signaling pathway (78.6% Fig. 4.2.A(II)), Dopaminergic synapse (76.7%), EPHA-mediated growth cone collapse (100%), CamPDE 1 activation (100%) and Adherens junction (83.3%) (see Table 4.2).

Figure 4.2: (Caption on next page.)

Figure 4.2: (Previous page.) **A. Activated Pathways in diffuse-myeloid synovia.**
(I) Activated network joint extracellular matrix genes for focal adhesion. **(II)** Gene
network of PPAR signaling pathway.
B. Activated Pathways in pauci-immune fibroid synovia. (I) Extracellular matrix
interaction clusters. **(II)** Genes of the Ras signaling pathway. **(III)** Clusters of Notch-,
Wnt- and TGF-beta signaling. **(IV)** Cytokine-cytokine interaction of pro-inflammatory
genes.

4.1.3.3 The pauci-immune fibroid pathotype

While few nodes were observed for the diffuse-myeloid pathotype and even fewer included unique functions, the pauci-immune fibroid pathotype has a more abundant network with 229 nodes and 299 edges (253 activators and 46 inhibitors). As in the previous analyses, the active links in common with the other pathotypes were discarded.

While there is substantial overlap with the previous two pathotypes regarding interactions related to the ECM pathway (Fig. 4.2.B(I)), the pauci-immune fibroid pathotype showed unique functions as evidenced, for example, by the presence of a wealth of growth factors and their receptors as part of the Ras signaling pathway (76.6% cluster genes, Fig. 4.2.B(II)).

Another node of major interest was found around the pro-inflammatory cytokine Interleukin (IL)-17D in a cluster of Cytokine-cytokine receptor interaction (100% cluster genes) shown in Figure 4.2.B(IV). Finally, additional clusters were Regulation of RUNX1 Expression and Activity (100%), AKT-mediated inactivation of FOXO1A (100%), mTOR signaling (100%), Adrenergic signaling in cardiomyocytes (81.8%) and Vascular smooth muscle contraction/Long-term depression (63.2%) (see Table 4.2).
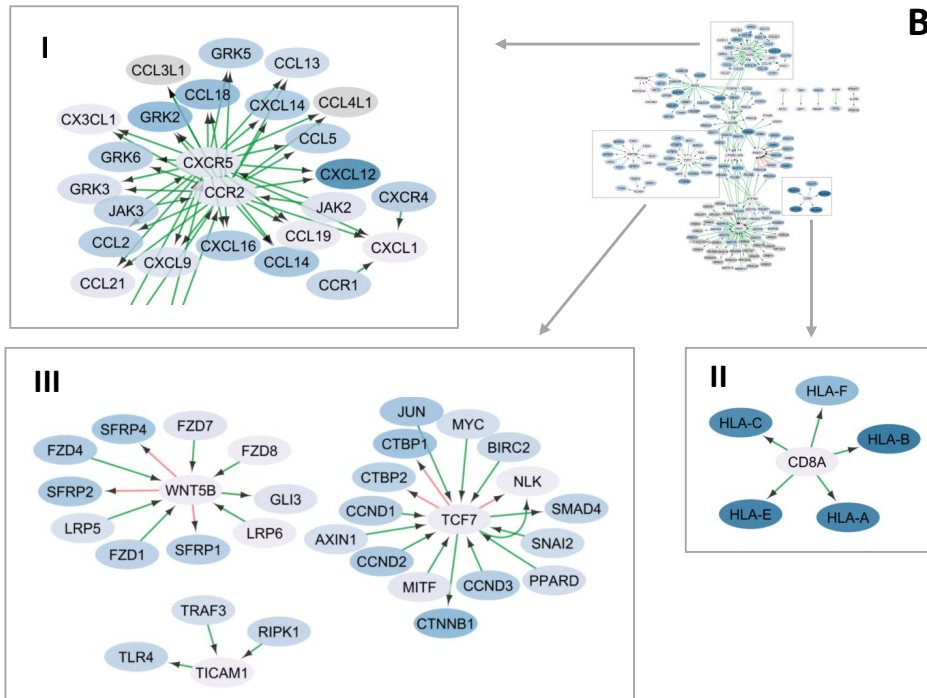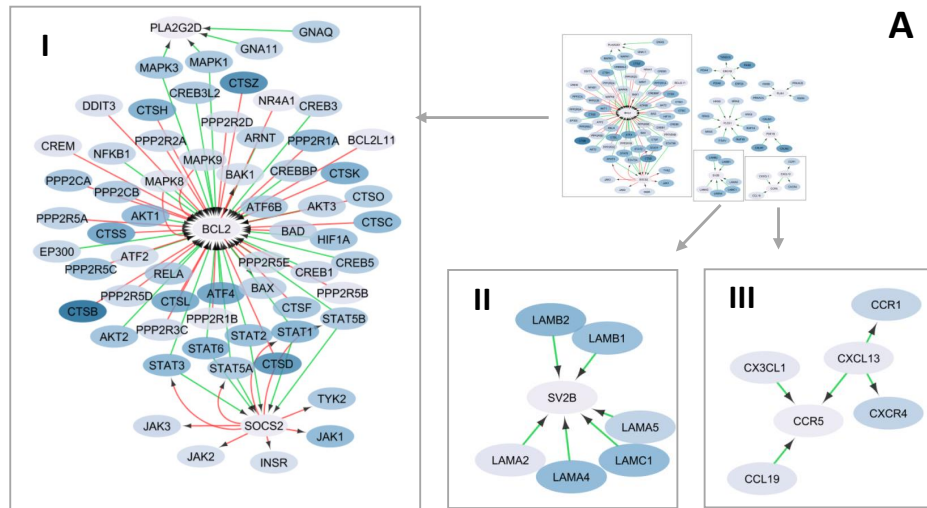
Figure 4.3: (Caption on next page.)

Figure 4.3: (Previous page.) **Interaction network of differentially expressed genes unique to Disease-modifying anti-rheumatic drug (DMARD) response.**
A. Clusters representing genes unique to DMARD responders **(I)** Cell survival genes as part of the PI3K-Akt signaling pathway in patients responding to DMARD treatment. **(II)** Cluster of ECM-receptor interaction genes is found in responders **(III)** Genes of chemokines and their receptors.
B Genes associated with DMARD non responders. **(I)** Chemokine signaling clusters linked to no response. **(II)** Antigen processing and presentation cluster. **(III)** Wnt signaling networks.

4.1.3.4 Baseline networks corresponding to response to csDMARD treatment

In the observations described above, three networks were built on the three histopathological categories (lympho-myeloid, diffuse-myeloid, and pauci-immune fibroid).

In the following section the extracted response and remission networks will be described. According to their responsiveness to six months of csDMARD treatments patients were categorised based on the DAS28 EULAR criteria. This divided the baseline cohort into three groups: good responders (n=26), moderate responders (n=32), and no responders (n=9) (see Table 4.3). Hence, three networks were extracted from the most actively interactions that are unique in the three groups.

In the interest of highlighting key mechanisms responsible for response/ remission, the moderate network is not described in this manuscript.

Patients who will respond well to csDMARD have few nodes of which the PI3K-Akt signaling pathway (53.1% cluster genes) cluster linked to B-cell lymphoma 2 (BCL2) is the most prominent one (Fig. 4.3.A (I)). Other clusters regarded signaling pathways as Signaling by MET (71.4%) and the Calmodulin (CaM) pathway (100%). The remaining two clusters were related to Triglyceride catabolism (100%) and disulfide oxidoreductase activity (100%) (see Table 4.2).

Non responders showed many active interactions across genes part of the Chemokine signaling pathway (100%, Fig 4.3.B (I)) around the node of

CCR2 and CXCR5. Two clusters showed Wnt signaling activity (Fig 4.3.B (III) 66.7% and 81.3%), while another small cluster showed Antigen processing and presentation activity (100%, Fig. 4.3.B (II)), as found in the lympho-myeloid group. The remaining clusters were associated with Olfactory transduction (64%), Cytosolic DNA-sensing pathway (83.3%), Human cytomegalovirus infection (61.5%) and Glioma (61.1%).

	Good (N=26)	Moderate (N=32)	None (N=9)	Total (N=67)
Pathotype				
- Lymphoid	15 (57.7%)	20 (62.5%)	6 (66.7%)	41 (61.2%)
- Myeloid	4 (15.4%)	8 (25.0%)	1 (11.1%)	13 (19.4%)
- Fibroid	7 (26.9%)	4 (12.5%)	2 (22.2%)	13 (19.4%)

Table 4.3: Distribution of synovial pathotypes, related to DMARD treatment response.

4.1.4 Discussion

4.1.4.1 Interactions between collagens, laminins and integrins recur in all classes

Focal adhesion or ECM interactions recur across all the networks with increased expression of collagens, laminins and integrins (Fig 4.1.C, 4.2.A(I), 4.2.B(I) and 4.3.A(II)). This suggests that a destructive process or at least and ECM remodelling process is already ongoing in RA early patients [69]. Nonetheless, a particular note must be reserved to the leucine-rich chondroadherin (CHAD) protein whose interaction with integrins is known to be involved in the pro-inflammatory and pro-catabolic response to cartilage matrix damage [70]. This interaction is significantly active for the lympho-myeloid and pauci-immune fibroid patients but absent in the diffuse-myeloid

class (see Fig. 4.4.A, B).

4.1.4.2 The lympho-myeloid pathotype shows molecular signature of leukocyte attracting chemokines

The lympho-myeloid network comprises of a wealth of chemokine genes (Fig. 4.1.A) including CCR1 and its ligands CCL5 and CCL14, CCR2 together with its ligand CCL2 and CCR7 and its ligands CCL19 and CCL21. Expression of these genes suggest a signature of leukocyte recruitment to the joints [71]. The lympho-myeloid pathotype is indeed classified as such because of the infiltration of lymphocytes. T cell genes are also prominent in the antigen processing and presentation cluster (Fig. 4.1.B). They include CD8A with different major histocompatibility complex (MHC) class I and T cell activation genes including CD247, ZAP70, CD28 and CD86. Alongside T-cell genes additional B cell recruitment is suggested by the presence of chemokine (C-X-C motif) ligand (CXCL13) in Fig. 4.1.A [72].

Other active pathways in the lympho-myeloid pathotype include NFKB and mammalian target of rapamycin (mTOR) signaling as part of chemokine and TNF signaling (Fig. 4.1.A and 4.1.D). A regulator of NFKB and mTOR signaling namely Inhibitor of nuclear factor kappa B kinase subunit epsilon (IKBKE) is found within an Interferon (IFN) signaling node where it also influences IRF genes (Fig. 4.1.E)[73]. Finally, a turnover of cells is suggested by clusters with cell death related genes for both the extrinsic and intrinsic apoptotic pathway with nodes in TRAF2, BAD, BAK, BCL2 and GZMB (Fig. 4.1.D and 4.1.F).

4.1.4.3 Macrophage activation and T cell monocytes in the diffuse-myeloid pathotype

The wealth of increased chemokine and cytokine expression detected in the lympho-myeloid environment (Fig. 4.1.A) was not present in diffuse-myeloid patients. However, CCR5 was also found to be increased and linked to CX3CL1 (data not shown). Activation of this interaction is evidence of chemoattractant activity for T cells and monocytes [74, 75], giving molecular confirmation of the histological definition.

Additionally, activation of genes from the PPAR signaling pathway can explain the presence of macrophages in the diffuse-myeloid pathotype [76]. The cluster is centered around PPARG and includes genes such as CYP27A1, ACSL1, LPL, SLC27A4 and its paralog SLC27A1, CD36, CPT1A, FABP4, SCD, PLIN1), ADIPOQ, ADIPOR1 and ADIPOR2.

4.1.4.4 A stimulating environment is found the synovium of the pauci-immune fibroid pathotype

Although the pauci-immune fibroid group demonstrates absence of immune cells infiltration, a wealth of growth factors still testifies dysregulation of the synovium tissue as stimulating environment.

In figure 4.2.B(II) we found a lot of growth factors with their receptors. As for example the epidermal growth factor (EGFR), fibroblast growth factors (FGF2, FGF7, FGFR1) that can be linked to the pathotype defining stromal cells, but also nerve growth factors (NGFR) that may explain angiogenesis in the joints, as well as insulin-like (IRS2, IGF1, INSR and IGF1R) and platelet-derived growth factors (PDGFRA, PDGFRB, PDGFC, PDGFD). Furthermore, these

nodes were associated with their signaling molecules through SRC, MAPK and PI3K.

Angiogenesis is also suggested by the presence of TNC [77] and ITGAV [78]. Figure 4.4.C shows the active interaction between LAMA3 and ITGAV that is not present in the other two histophatological classes.

Finally, we found activation of the WNT signaling pathway as consequence of fibroblast presence in the synovium tissue [79].

4.1.4.5 Good responders to csDMARD show enhanced apoptosis at baseline and expression of CXCL13

The network of active interactions characterising the good responders class presents few clusters offering key suggestions.

First, the whole network is dominated by a big cluster centered around BCL2 (Fig. 4.3.A(I)). This gene was found to be highly expressed in the synovial tissue of RA patients, particularly in CD68- synovial fibroblasts[80]. However, its interaction with pro-apoptotic genes from the same family (BAX, BAD, BAK1, BCL2L11) and many cathepsins needed for caspase activation (CTSB, CTSC, CTSD, CTSF, CTSK, CTSL, CTSO, CTSS, CTSZ) confirms previous hypotheses of beneficial effects of enhanced apoptosis in RA [81]. As shown in Figure 4.1.F, we observed these links to be equivalently active in the lympho-myeloid pathotype. This overlap can be thought as another potential explanation of lympho-myeloidsâĂŹ higher response to DMARDs compared to pauci-immune fibroids response [67]. In fact, activation of synovial fibroblasts has been correlated with apoptosis activity reduction [82, 83].

Furthermore, the same cluster shows centrality of another interesting gene,

part of the suppressor of cytokine signaling family, namely SOCS2. Its interaction with JAK-STAT genes is evidence of an immunosuppression process that is not active in the non responder class (Fig. 4.4.D as an example)[84].

Another minor cluster belonging to the good responders network detected chemokine receptors binding chemokines (Fig. 4.3.A (III)). Genes of this family were also found in the non responders network but with greater abundance (Fig. 4.3.B (I)). Most notably, the good responders class showed stronger activation of CXCL13 (Fig. 4.4.E) that is believed to be marker of early joint inflammation still responsive to csDMARD treatment [85].

4.1.4.6 Non responders to csDMARD show expression of CXCR5 and activation of the WNT signaling pathway

The non responders network showed leukocyte attracting chemokines (Fig. 4.3.B (I)) as in the lympho-myeloid pathotype. Of note, this class demonstrated differential activation of CXCR5 compared to the good responder counterpart (Fig. 4.4.F). Expression of this gene has been considered essential for the induction of RA since CXCR5 deficiency completely prevented RA development in mice [86].

Additionally, the non responders showed similarities to the pauci-immune fibroid class as we obtained two Wnt signaling clusters (Fig 4.3.B (III)). However, they differed in the containing genes as for example only WNT5B but not WNT11 was included and also Dishevelled proteins were missing.

Finally, a small cluster containing antigen processing and presentation genes (Fig. 4.3.B (II)) seem to suggest that in this class of patients the state of inflammation is already high.

Other interesting considerations can be done on the differential role of NOS3 in good and none responders. NOS3 contributes in the regulation of inflammation, endothelial functions, and bone remodeling process [87]. However its role is still unclear in many aspects. According to our study NOS3 acts as activator of AKT3 and inhibitor of CAMK1 (Fig. 4.4.G, 4.4.H). AKT is known to be a regulator of TNF-alpha-mediated apoptosis in RA synovial fibroblasts [88], while CAMK1 mediates the macrophage inflammatory response to sepsis [89].

4.1.5   Pathways Overview

In the above sections we provided a detailed description of gene interactions characterising three distinct RA pathotypes. By means of Pat-BINE, we looked inside pathways and explored those links actually active into each group.
In this final section we will show a general overview of the key disregulated pathways identified above. We compared their overall alterations across pathotypes using the the tool described in Chapter 3, Section 3.2.
In Figure 4.5 pathways with higher up-regulation in the lympho-myeloid group are shown.
Figure 4.6.A shows higher up-regulation of the Peroxisome proliferator-activated receptor (PPAR) signaling pathway in the diffuse-myloid group. Finally, the extracellular matrix (ECM) receptor pathway, up-regulated among the three groups, looks mostly altered in the pauci-immune fibroid group (Figure 4.6.B). All the showed box plots obtained a Kruskal-Wallis test p-value less then 0.05.

4.1.6 Section Summary

In this Section we applied a novel Patients' Biological Interactions Networks Extractor (Pat-BINE) to a large cohort of early rheumatoid arthritis patients. To systematically analise the extracted networks we utilised the ClueGO Cytoscape plugin and obtained GO terms associated to clusters inside networks. Finally, a new MITHrIL feature was applied to the cohort to generate an overview of key altered pathways.

Results of the Pat-BINE analysis applied to three histopathological categories showed a coherent genetic signature of the histological description. This justified the use of the proposed tool to assess responses/remission marks and look for correlations between pathotypes and response classes. To this aim, characteristic networks of gene interactions were extracted for both good and non responders to csDMARD therapy.

Patients who well responded showed enhanced apoptosis at baseline and expression of CXCL13. Furthermore, there was sign of an ongoing immunosuppression process through Suppressor Of Cytokine Signaling 2 (SOCS2). On the other hand, the non responders group showed expression of CXCR5 that is believed to play a key role in RA development and activation of the WNT signaling pathway, observed also in the pauci-immune fibroid pathotype.
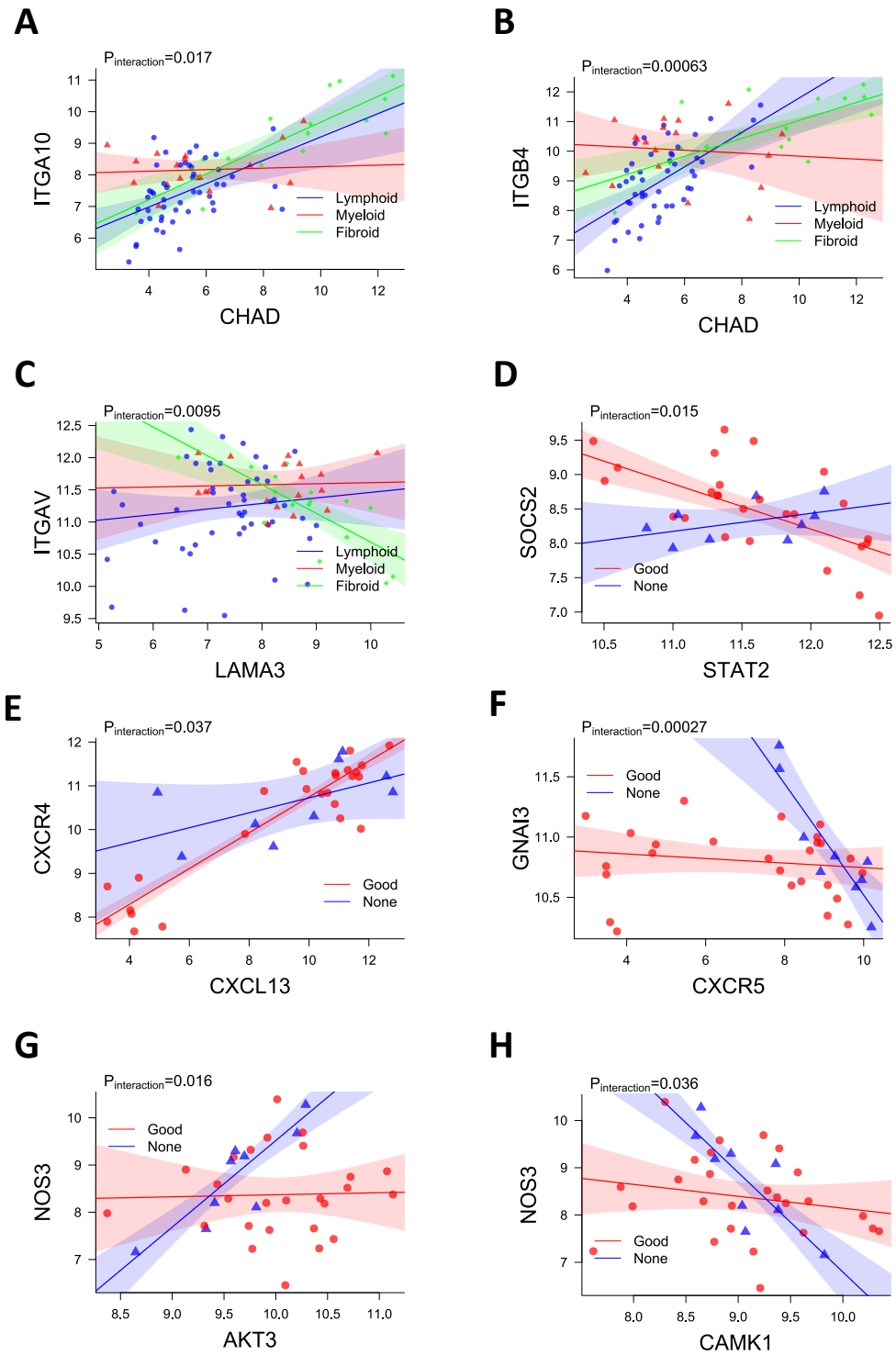
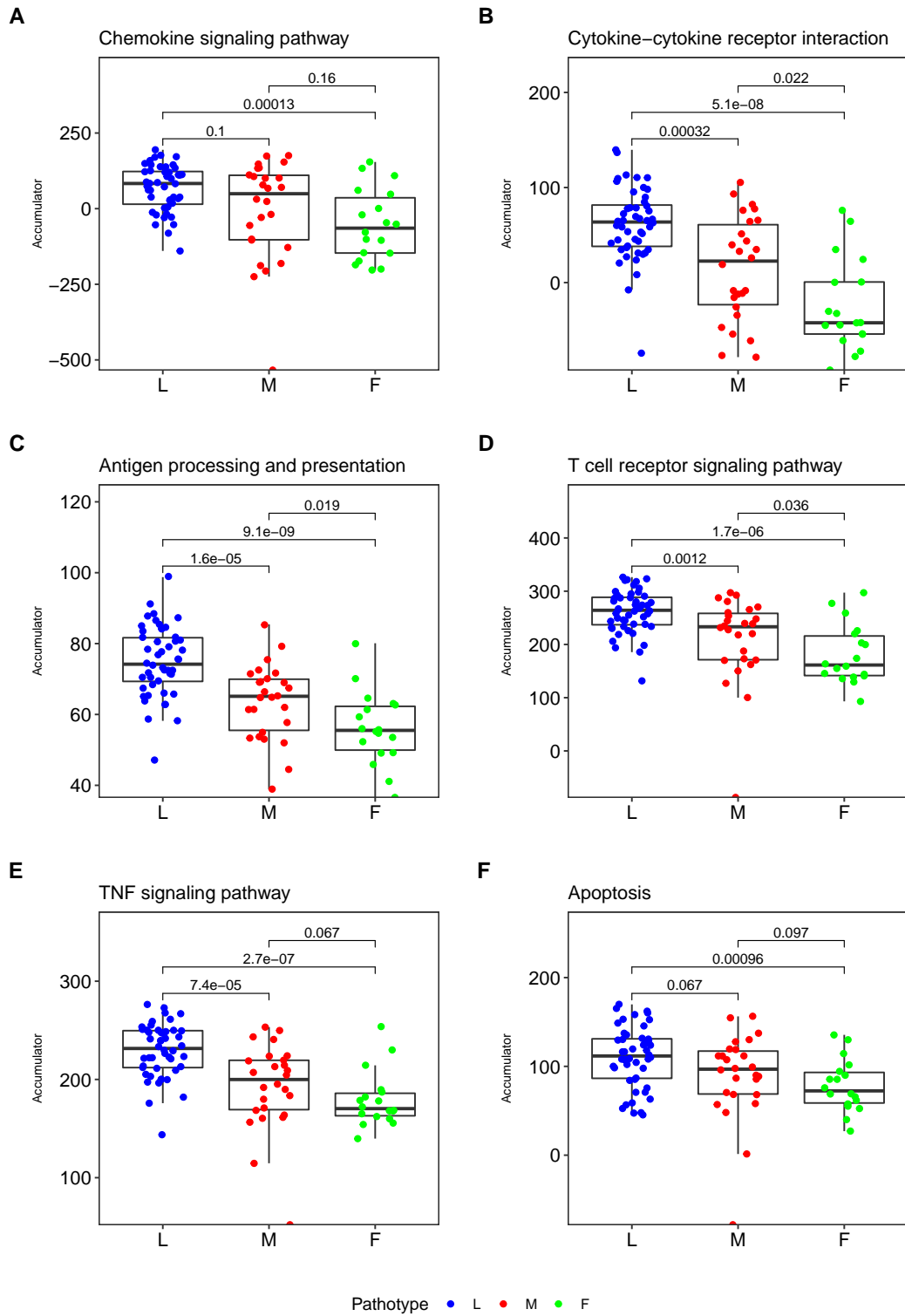**Figure 4.4:** Significant gene-gene interactions in PEAC.

Figure 4.5: Disregulated pathways across RA pathotypes (Part I).
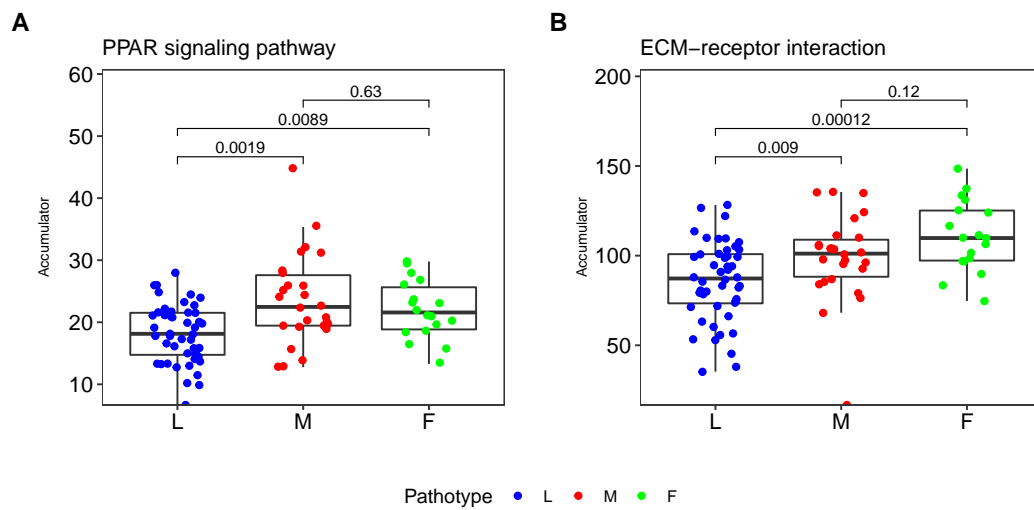*L = Lympho-myeloid, M = diffuse-myeloid, F = pauci-immune fibroid*

Figure 4.6: Disregulated pathways across RA pathotypes (Part II).
L = Lympho-myeloid, M = diffuse-myeloid, F = pauci-immune fibroid

4.2 Trial of anti-B-cell Therapy in patients with primary Sjögren's Syndrome

A transcriptomic study on a B-cell depletion therapy in Sjögren's syndrome patients

Primary Sjögren's syndrome (pSS) is a chronic, systemic autoimmune disease characterised by the impaired secretory function of the exocrine glands, in particular the salivary and lachrymal glands [90].

Despite the progression in understanding the pathogenic mechanism occurring in pSS, treatment options are currently symptom dependent. The current therapies focus on relieving symptoms, rather than changing the actual course of disease.

Targeting B-cells is certainly a rationale therapeutic approach in diseases like pSS, where B-cell hyperactivity is believed to be an essential pathogenic event [91, 92].

The TRial of Anti-B-Cell Therapy In patients with primary Sjögren's Syndrome (TRACTISS) is the biggest UK multi-centre, double-blind, randomised, controlled, parallel group trial of 110 patients with Sjögren's syndrome [6]. The main goal of the clinical trial is to investigate whether Rituximab (MabThera®/ Rituxan®) leads to an improvement in patient symptoms and a reduction in disease damage and activity [93, 94]. Rituximab is an immuno-modulatory agent depleting selectively B-cells, the antibody producing lymphocytes [95]. The patient cohort includes two arms of treatment: one treated with Rituximab and the other treated with placebo. Salivary glands biopsies and matched peripheral blood were collected from both treatment (Rituximab) and control (placebo) group of patients. RNA-seq analysis represents the first longitudinal

study (three sequential biopsies over a year time) of salivary glands biopsies in pSS.

The transcriptomic analysis in the placebo arm informs on the disease evolution, identifying the inflammatory mediators driving peripheral B-cell hyperactivity and salivary gland autoimmune response. While, the same analysis in the Rituximab arm, characterises the effects of the B-cell depletion therapy. The analyses described in this chapter are restricted to only biopsies from salivary glands. Peripheral blood samples will be analysed in a future work.

4.2.1 Patient Cohort

Transcriptomic analysis of salivary gland tissues was performed on 29 patients whose biopsies were collected at three time points (0, 16 and 48 weeks). Some patient did not underwent biopsy at all timepoints, so the total number of samples is 51. After quality checks two samples were excluded (see Section 4.2.2), the remaining 49 samples are distributed across timepoints as follow: 23 at week 0, 15 at week 16, and 14 at Week 48 (see Table 4.4).

4.2.2 Quality Checks and Preliminary Analysis

Prior to start the analysis, integrity and quality checks were performed on data.

First of all, quality assessment parameters, such as RIN score and DV 200, were verified according to best practice.

The RNA Integrity Number (RIN) is the result of an automated algorithm based on a combination of different features contributing to RNA integrity

	Placebo (N=15)	RTX (N=8)	Total (N=23)	p value
Age				0.263
- <65	11 (73.3%)	4 (50.0%)	15 (65.2%)	
- ≥ 65	4 (26.7%)	4 (50.0%)	8 (34.8%)	
Gender				0.433
- Female	11 (73.3%)	7 (87.5%)	18 (78.3%)	
- Male	4 (26.7%)	1 (12.5%)	5 (21.7%)	
Disease duration (months)				0.482
- <10	13 (86.7%)	6 (75.0%)	19 (82.6%)	
- ≥ 10	2 (13.3%)	2 (25.0%)	4 (17.4%)	
ESSDAI	6.1 (5.2)	3.0 (2.3)	5.0 (4.6)	0.126
IgG	19.0 (10.5)	16.5 (7.1)	18.2 (9.4)	0.577
IgA	3.5 (2.4)	2.3 (0.7)	3.1 (2.1)	0.241
IgM	1.4 (0.6)	1.3 (0.6)	1.3 (0.6)	0.963
C3	1.4 (0.4)	1.2 (0.3)	1.3 (0.4)	0.264
C4	0.3 (0.1)	0.2 (0.1)	0.2 (0.1)	0.078
RHF	112.3 (158.4)	25.4 (20.1)	81.9 (133.4)	0.171

Table 4.4: Baseline demographics of treatment-naïve pSS patients recruited in this study.

[96]. Its value can vary from 1 to 10, with values >6 assessing absence of degradation and good quality. All the samples belonging to the cohort satisfied this requirement.

On the other hand, DV200 evaluates the percentage of fragments of >200 nucleotides [97]. It can vary from 0% to 100% and was found to be less then 70 in one sample, indicating short fragment band.

Additionally, the same sample turned to be an outlier in the Principal Component Analysis (PCA) done on the whole dataset (Figure 4.7). Because of this, the sample has been removed from the cohort.

Additional PCAs were performed on each timepoint subset. They revealed

another sample to be outlier among all week 16 samples (Figure 4.8). The latter was also found outlier in the global PCA, when looking at the fourth Principal Component (PC4) (Figure 4.9). While further investigating on other possible degradation issues, we found the quantity of extracted RNA was poor for the mRNA sequencing. This led this second sample to be excluded from the subsequent analysis.

After outliers removal, the new global PCA (including all timepoints) looks neat and not clustered on timepoints (Figure 4.10).

It rather seems to cluster on gender, with the exception of two male samples (Figure 4.11).
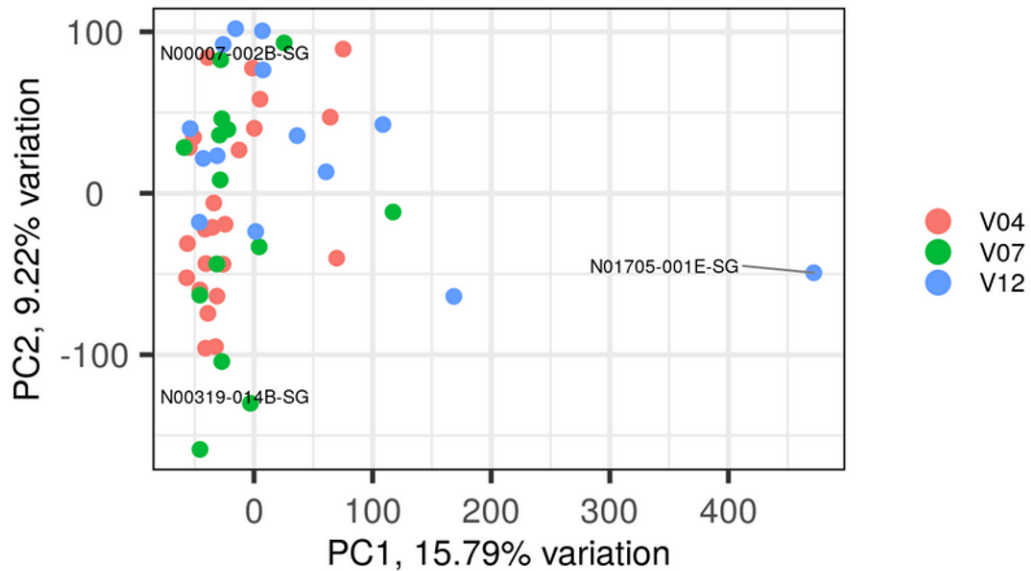

Figure 4.7: Principal Component Analysis of the whole cohort. Color scheme on timepoints.
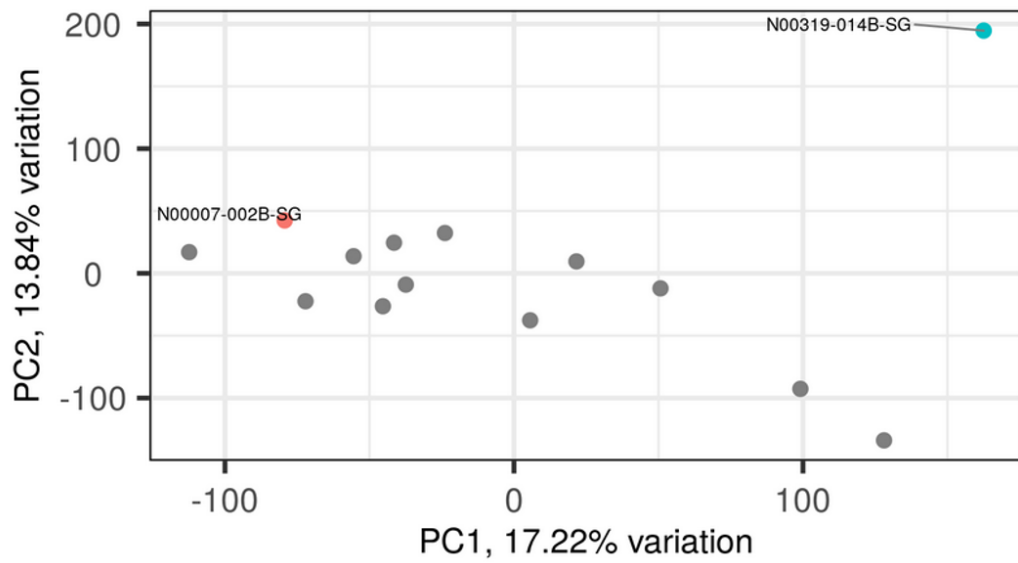
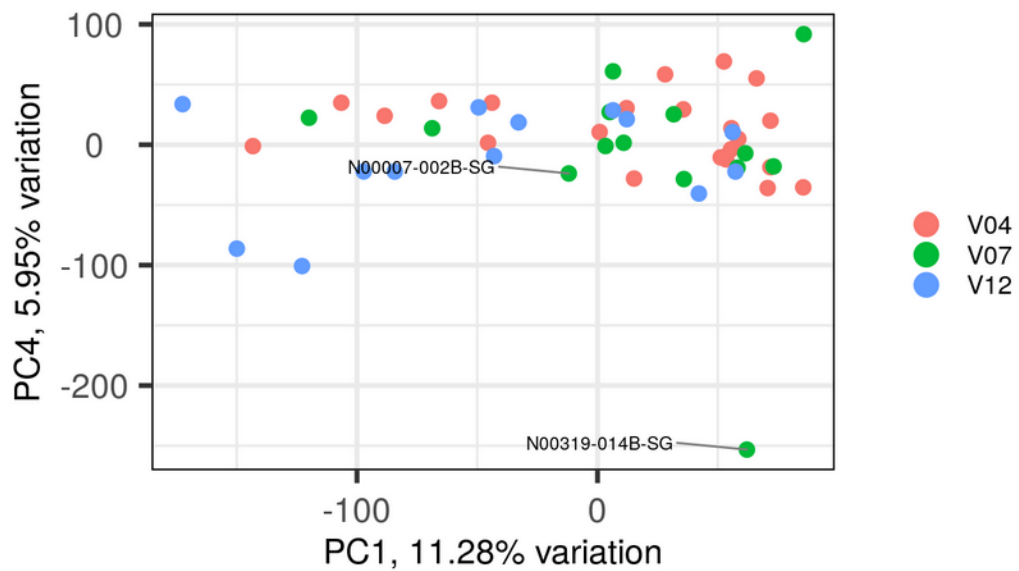**Figure 4.8:** Principal Component Analysis of week 16 samples.

Figure 4.9: Principal Component Analysis of the whole cohort, first and fourth Pricipal Components shown. Color scheme on timepoints.
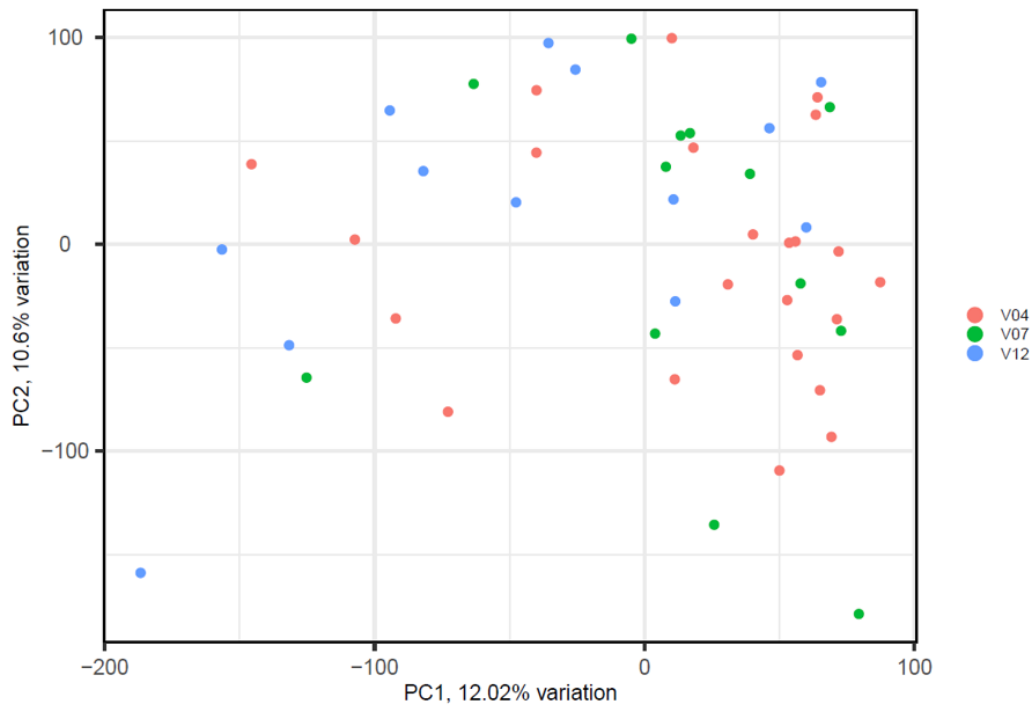
Figure 4.10: Principal Component Analysis of the whole cohort, after outliers removal. Color scheme on timepoints.

4.2.3 Rituximab Vs Placebo

To assess the effect of B-cell depletion on pSS patients, we used the *DESeq2* software [59] to perform a Differentially Expressed Gene (DEG) analysis on Rituximab Vs placebo biopsies after 48 weeks of treatment. Q-values were derived from the estimated p-values by means of *qvalue* software [98].

The DEG analysis highlighted 157 significantly differentially expressed genes (i.e. q-value < 0.05) that are shown in Figure 4.12.

Placebo samples - compared to RTX - showed up-regulation of immune cell genes, mainly related to B- and T-cell activation (CD79B, PAX5, CD72, CD3D) and lymphoid neogenesis (CXCL13, CXCR5, CCL19,LTB) known to be
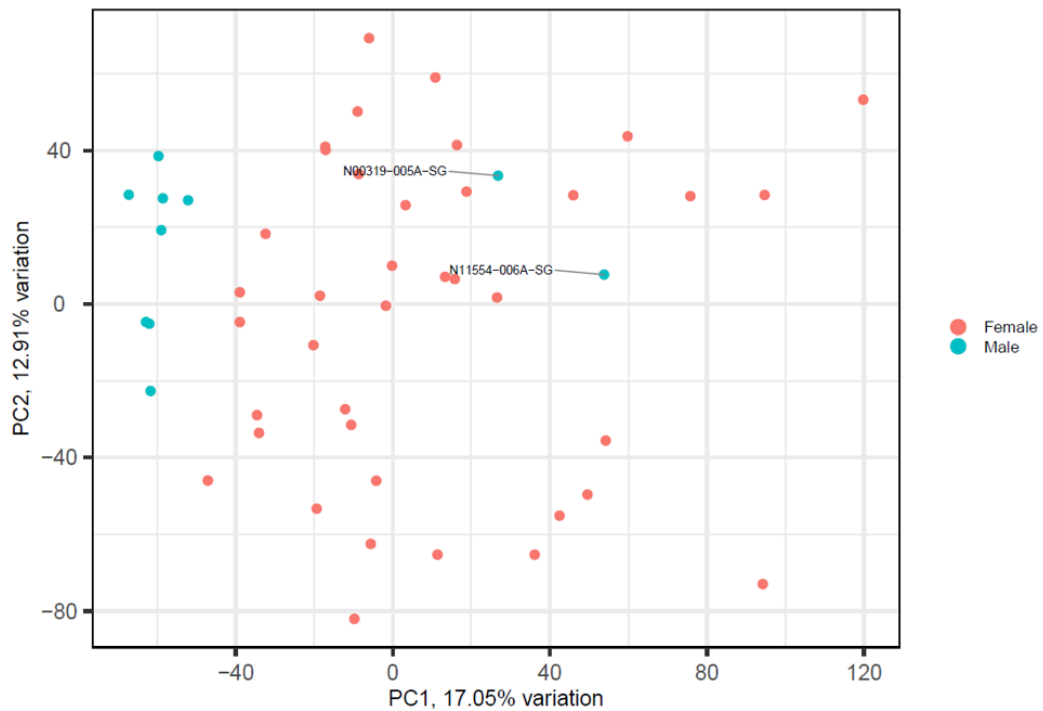
Figure 4.11: Principal Component Analysis of the whole cohort, after outliers removal. Color scheme on gender.

pathogenic events in the salivary glands of pSS patients.

The same set of genes (157 DEGs) were given in input to MITHrIL [49] to extract the related disregulated pathways (see Figure 4.15.

This pathway analysis pointed out a global down-regulation of inflammatory pathways suggesting a good impact of Rituximab in the salivary gland inflammation.

The same group of genes was used again to drawn the heatmap shown in Figure 4.13.

Using different color hues, the heatmap allows to look at the magnitude of gene alterations for each sample (in columns). Moreover, samples and genes

are sorted according a clustering algorithm that tends to bring similar samples/genes closer to each other. In this particular implementation, rows (genes) were clustered using the complete linkage method [99] based on euclidean distance, while in columns (samples) the Ward's hierachical clustering [100] was used. In this case, based on the gene expression of the DEGs, samples are clearly segregated by treatment (see *Treatment* track on top) and, most interestingly, the placebo group is in turn segregated in two clusters matching with gender (see *Gender* track on top). This suggests that the up-regulation of B-cell-lineage genes (such as PAX5, chemokine receptors and ligands involved in B- and T-cell recruitment) and ectopic lymphoid-related genes (such as LTB, CXCL13, CXCR5, CCL19) was mainly driven by the female subjects.

Moreover, the histological characterization of CD20+ B- and CD3+ T-cell infiltration over a year follow-up in the placebo group confirms the worsening of inflammatory infiltration in female, but not in the male subjects recruited in the study (Figure 4.14). In line with the higher prevalence of pSS in females compared to males with a ratio of 9 to 1, these results show for the first time how female and male salivary glands are characterized by a complete different transcriptomic profile with an up-regulation of immune cells for females.

4.2.4   Time Course Analysis

In the previous section we compared RNA-seq of placebo and Rituximab treated patients collected at the same timepoint (i.e. after 48 weeks of treatment).

In this section we will look how lymphoid neogenesis genes change over time
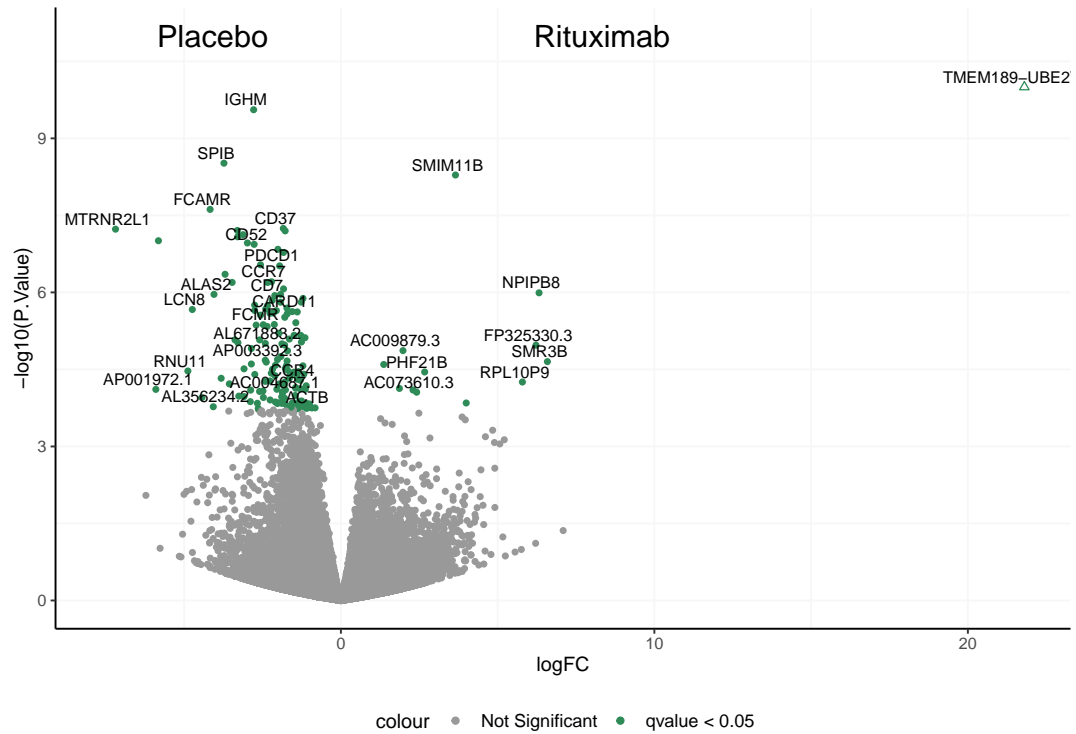
Figure 4.12: Volcano plot showing the differentially expressed genes between Rituximab and placebo treated samples at week 48.

(i.e. across three timepoints) in the salivary glands tissue of both placebo and Rituximab group.

To this aim, gene expression level distributions over time of chemokine receptor and ligands involved in the B- and T-cell recruitment such as CCL21 and CXCR4 have been reported for both placebo and Rituximab samples (Figure 4.16). Those genes are both down-regulated after two doses of Rituximab. The same occurs for ICOS and PDCD1 (Figure 4.17), involved in the B- and T-cell organisation. Similarly, the B-cell lineage PAX5 gene is lowered after the treatment with Rituximab (Figure 4.18).

These findings might explain the effect of Rituximab in preventing newly
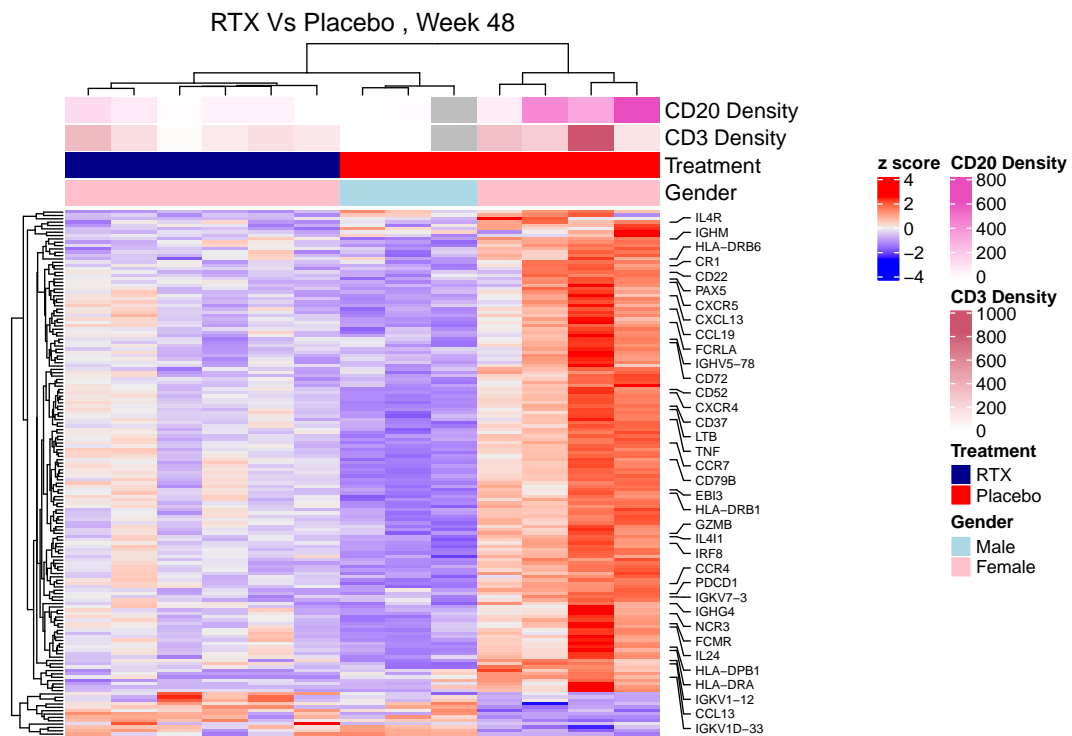
Figure 4.13: Heatmap showing expression levels of genes of samples at 48 weeks. Only the previously identified DEGs are shown.

immune cell infiltration and the organization of ectopic germinal centres, observed from the longitudinal histological analysis of placebo group, reported in Figure 4.14.

4.2.5 Section Summary

In this Section a transcriptomic analysis on the TRial of Anti-B-Cell Therapy In patients with primary Sjögren's Syndrome (TRACTISS) has been reported. For the first time, a longitudinal analysis has been conducted on salivary gland tissues from both Rituximab and placebo treated patients over a year time. This revealed an important role of Rituximab in preventing the salivary gland

inflammation worsening. Moreover, on the placebo arm, the RNA-seq analysis has revealed clear segregation between males and females transcriptomic profiles, showing a greater inflammatory activity in women (in consistence with disease incidence).
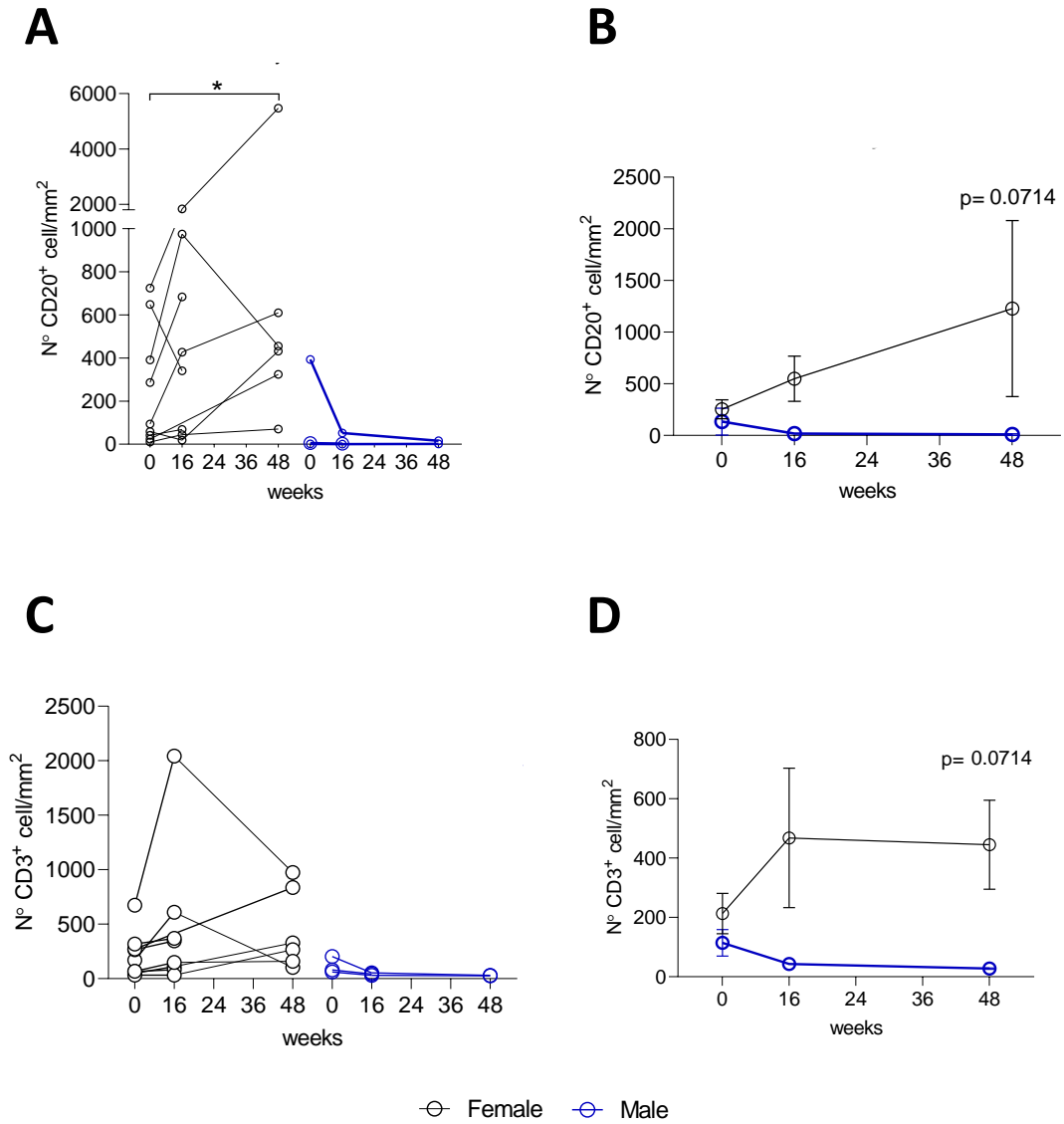
Figure 4.14: Histological characterisation of CD20+ B- and CD3+ T-cell infiltration over 48 weeks in placebo patients.
(A, C) Levels of B- and T-cells density over time per patient. **(B, D)** Average levels of B- and T-cells density over time.
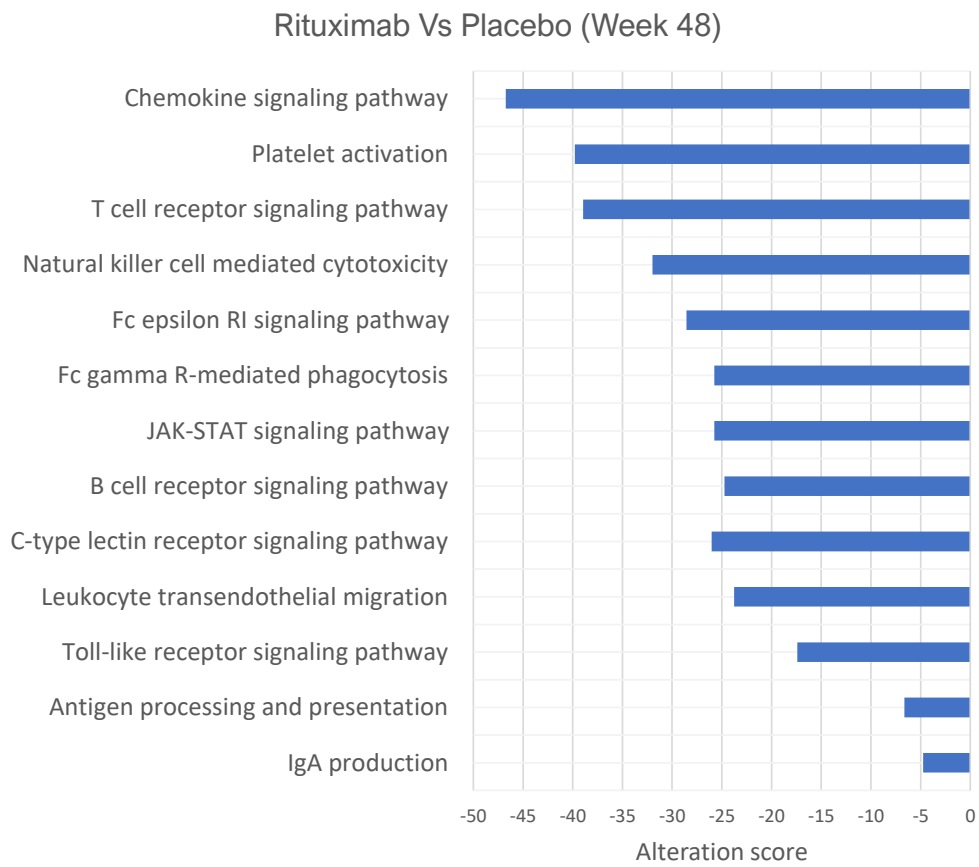
Figure 4.15: Disregulated pathways of the Rituximab group compared to placebo. Only immune system related pathways are shown.
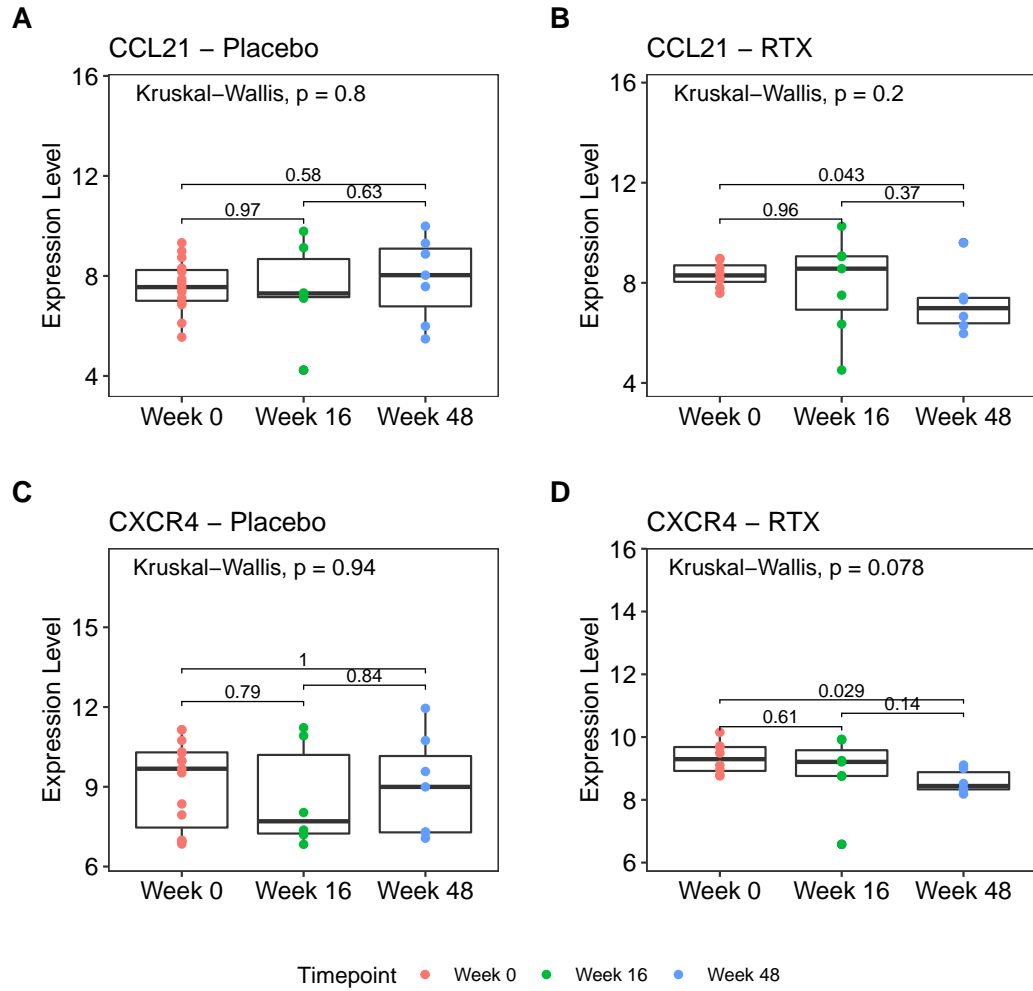
Figure 4.16: Expression level distributions of CCL21 and CXCR4 over time in the placebo group (A, C) and Rituximab group (B, D).
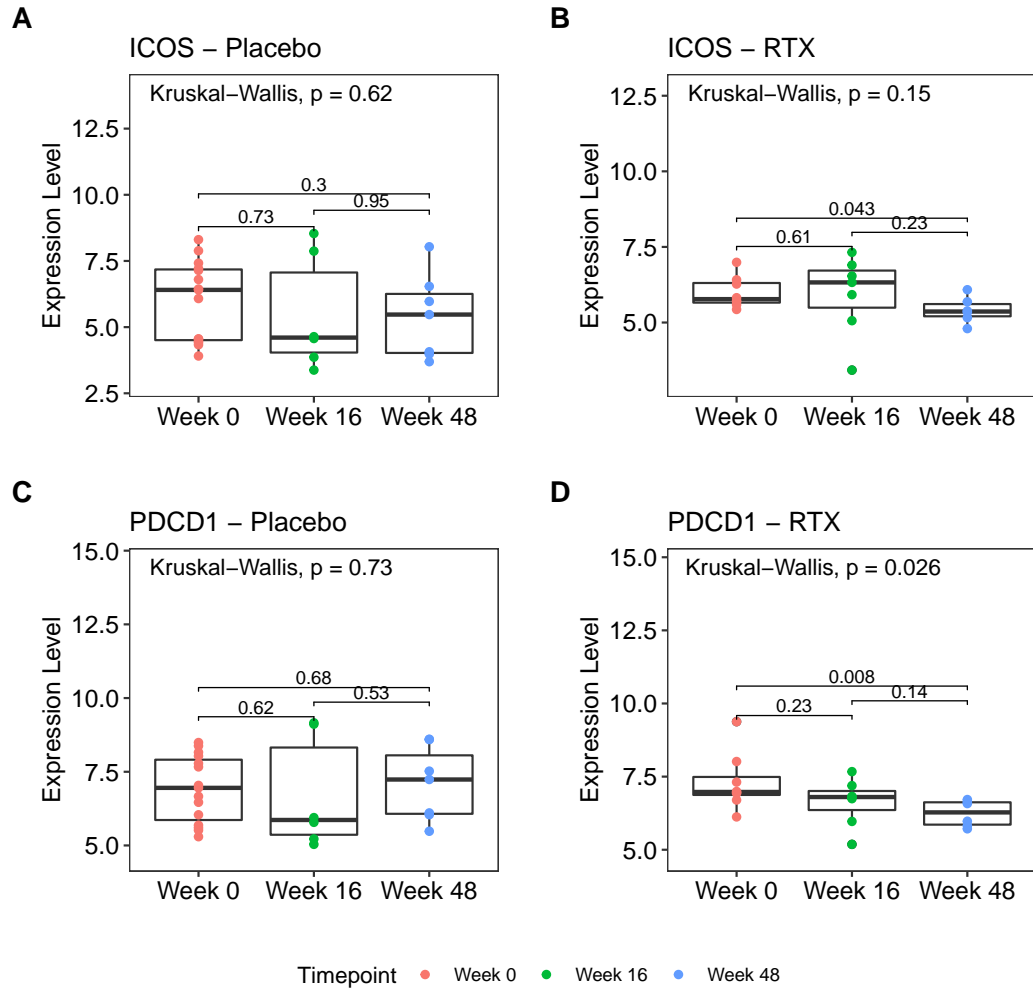
Figure 4.17: Expression level distributions of ICOS and PDCD1 over time in the placebo group (A, C) and Rituximab group (B, D).
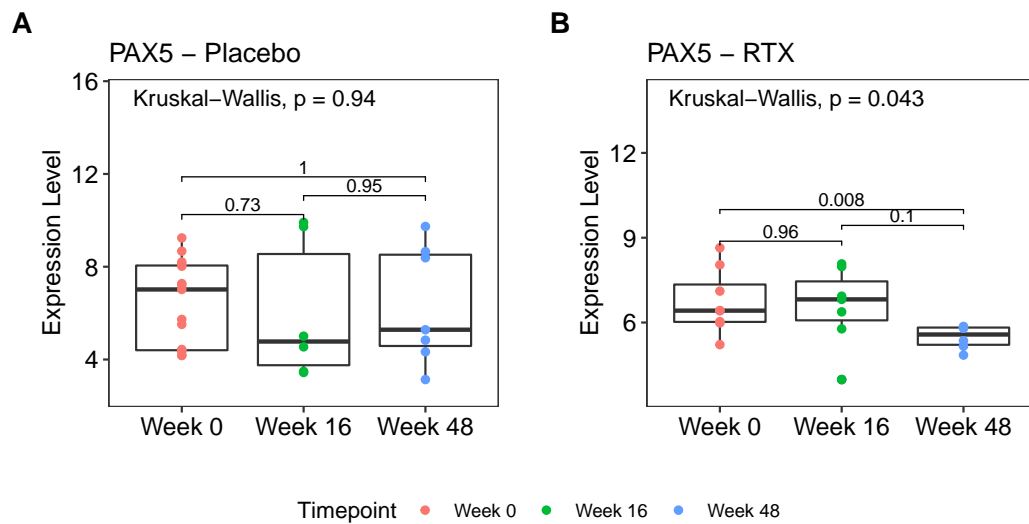
Figure 4.18: Expression level distribution of PAX5 over time in the placebo group (A) and Rituximab group (B).

Chapter 5

Conclusion

In this thesis engineering and analytical tools have been discussed to support a paradigm shift towards precision medicine.

Chapter 2 concerns safety and reliability of communications in intra-body scenarios, where implanted sensors are deployed for clinical parameters monitoring, allowing adjustment of ongoing therapies.

The use of ultrasounds is envisaged and a testbed has been developed to emulate human body propagation features. An impulse response analysis of this channel has been conducted and other key metrics of channel behavior were also evaluated as a function of different parameters such as frequency, channel composition and distance.

Chapter 3 and 4 focus on an another stage of the precision medicine process, where genomic signatures are identified to stratify patients.

In Chapter 3 a novel methodology called Pat-BINE has been presented. This allows to bring out characterising gene interactions activated in user-defined groups of patients. In the same chapter, for a higher level view, another feature to visualise altered pathways over multiple groups is discussed.

Finally, in Chapter 4 analysis and results from two clinical trials are reviewed. In the first section, Pat-BINE has been used to assess responses/remission marks to csDMARDs therapies in rheumatoid arthritis patients. In the second section, a longitudinal transcriptomic analysis in Siögren's Syndrome affected patients is conducted to determine the effect of a B-cell depleting therapy.

References

1. Jameson, J. L. & Longo, D. L. Precision medicine — personalized, problematic, and promising. *Obstetrical & gynecological survey* **70,** 612–614 (2015).

2. Ashley, E. A. Towards precision medicine. *Nature Reviews Genetics* **17,** 507 (2016).

3. Scarpato, N., Pieroni, A., Di Nunzio, L. & Fallucchi, F. E-health-IoT universe: A review. *management* **21,** 46 (2017).

4. Cheol Jeong, I., Bychkov, D. & Searson, P. C. Wearable devices for precision medicine and health state monitoring. *IEEE Transactions on Biomedical Engineering* **66,** 1242–1258 (2018).

5. Lewis, M. J., Barnes, M. R., Blighe, K., Goldmann, K., Rana, S., Hackney, J. A., Ramamoorthi, N., John, C. R., Watson, D. S., Kummerfeld, S. K., Hands, R., Riahi, S., Rocher-Ros, V., Rivellese, F., Humby, F., Kelly, S., Bombardieri, M., Ng, N., DiCicco, M., van der Heijde, D., Landewé, R., van der Helm-van Mil, A., Cauli, A., McInnes, I. B., Buckley, C. D., Choy, E., Taylor, P. C., Townsend, M. J. & Pitzalis, C. Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Reports* **28,** 2455–2470.e5. ISSN: 22111247 (2019).

6. Brown, S., Coy, N. N., Pitzalis, C., Emery, P., Pavitt, S., Gray, J., Hulme, C., Hall, F., Busch, R., Smith, P., *et al.* The TRACTISS Protocol: a randomised double blind placebo controlled clinical TRial of Anti-B-Cell Therapy In patients with primary Sjögren's Syndrome. *BMC musculoskeletal disorders* **15,** 21 (2014).

7. Sciacca, E. C. & Galluccio, L. Impulse response analysis of an ultrasonic human body channel. *Computer Networks* **171,** 107149 (2020).

8. Cheung, A. Y. & Neyzari, A. Deep local hyperthermia for cancer therapy: external electromagnetic and ultrasound techniques. *Cancer Research* **44,** 4736s–4744s (1984).

9. Rivet, F., Redois, S. & Deval, Y. *Characterization of ultrasonic wave propagation for intra-body communication* in (2016).

10. Santagati, G. E. & Melodia, T. *An implantable low-power ultrasonic platform for the Internet of Medical Things* in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications* (2017), 1–9.

11. Miller, D. L., Smith, N. B., Bailey, M. R., Czarnota, G. J., Hynynen, K., Makin, I. R. S. & of the American Institute of Ultrasound in Medicine, B. C. Overview of therapeutic ultrasound applications and safety considerations. *Journal of ultrasound in medicine* **31,** 623–634 (2012).

12. Ifantis, A. & Kalis, A. *On the use of ultrasonic communications in biosensor networks* in *2008 8th IEEE International Conference on BioInformatics and BioEngineering* (2008), 1–6.

13. Galluccio, L., Melodia, T., Palazzo, S. & Santagati, G. E. *Challenges and implications of using ultrasonic communications in intra-body area networks* in *2012 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS)* (2012), 182–189.

14. Santagati, G. E., Melodia, T., Galluccio, L. & Palazzo, S. Ultrasonic networking for e-health applications. *IEEE Wireless Communications* **20,** 74–81 (2013).

15. Santagati, G. E. & Melodia, T. *Sonar inside your body: Prototyping ultrasonic intra-body sensor networks* in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (2014), 2679–2687.

16. Santagati, G. E. & Melodia, T. Opto-ultrasonic communications for wireless intra-body nanonetworks. *Nano Communication Networks* **5,** 3–14 (2014).

17. Dressler, F. & Fischer, S. Connecting in-body nano communication with body area networks: Challenges and opportunities of the Internet of Nano Things. *Nano Communication Networks* **6,** 29–38 (2015).

18. Charthad, J., Weber, M. J., Chang, T. C. & Arbabian, A. A mm-sized implantable medical device (IMD) with ultrasonic power transfer and a hybrid bi-directional data link. *IEEE Journal of solid-state circuits* **50,** 1741–1753 (2015).

19. Laqua, D, Sühn, T, Kring, K, Albrecht, K & Husar, P. Ultrasound communication for intelligent implants. *Biomed Tech* **59** (2014).

20. Singer, A., Oelze, M. & Podkowa, A. *Mbps experimental acoustic through-tissue communications: MEAT-COMMS* in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (2016), 1–4.

21. Li, M. & Kim, Y. T. Feasibility Analysis on the Use of Ultrasonic Communications for Body Sensor Networks. *Sensors* **18,** 4496 (2018).

22. Galluccio, L., Milardo, S. & Sciacca, E. *A feasibility analysis on the use of ultrasonic multihop communications for e-health applications* in *2017 IEEE International Conference on Communications (ICC)* (2017), 1–6.

23. Galluccio, L., Milardo, S. & Sciacca, E. *Demo abstract an ultrasonic intra body area network for ehealth applications* in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (2017), 982–983.

24. Guan, Z., Santagati, G. E. & Melodia, T. *Ultrasonic intra-body networking: Interference modeling, stochastic channel access and rate control* in *2015 IEEE Conference on Computer Communications (INFOCOM)* (2015), 2425–2433.

25. Guan, Z., Santagati, G. E. & Melodia, T. Distributed algorithms for joint channel access and rate control in ultrasonic intra-body networks. *IEEE/ACM Transactions on Networking* **24,** 3109–3122 (2016).

26. Santagati, G. E., Melodia, T., Galluccio, L. & Palazzo, S. Medium access control and rate adaptation for ultrasonic intrabody sensor networks. *IEEE/ACM Transactions on Networking* **23,** 1121–1134 (2014).

27. Demirors, E., Alba, G., Santagati, G. E. & Melodia, T. *High data rate ultrasonic communications for wireless intra-body networks* in *2016 IEEE International Symposium on Local and Metropolitan Area Networks (LAN-MAN)* (2016), 1–6.

28. Pratama, M. H. B., Munandar, A., Mujib, K., Wicaksono, E. S. & Zahra, A. A. *Implementation of ultrasonic communication for wireless body area network using amplitude shift keying modulation* in *2016 IEEE Region 10 Conference (TENCON)* (2016), 3790–3793.

29. Canovas-Carrasco, S., Garcia-Sanchez, A.-J. & Garcia-Haro, J. A nanoscale communication network scheme and energy model for a human hand scenario. *Nano communication networks* **15,** 17–27 (2018).

30. Wild, G. & Hinckley, S. *Wireless acoustic communications for in-vivo biomedical device monitoring* in *Biomedical Applications of Micro-and Nanoengineering IV and Complex Systems* **7270** (2008), 72700T.

31. Jiang, W., Bos, T., Dehaene, W., Verhelst, M. & D'hooge, J. *Modelling of Channels for Intra-Corporal Ultrasound Communication* in *2018 IEEE International Ultrasonics Symposium (IUS)* (2018), 1–4.

32. Bos, T., Jiang, W., D'hooge, J., Verhelst, M. & Dehaene, W. Enabling ultrasound in-body communication: FIR channel models and QAM experiments. *IEEE transactions on biomedical circuits and systems* **13,** 135–144 (2018).

33. Hogg, T. & Freitas Jr, R. A. Acoustic communication for medical nanorobots. *Nano Communication Networks* **3,** 83–102 (2012).

34. slideshare.net. *Longitudinal, Shear, Rayleigh and Love Waves* https://www.slideshare.net/mboria/muravin-acoustic-emission-wave-propagation-and-source-location. 2009.

35. OLYMPUS CORPORATION. *Ultrasonic Tranducers Olympus V326-SU.* <https://www.olympus-ims.com/it/ultrasonic-transducers/immersion/{\%}5C{\#}!cms[focus]=descU8423054>.

36. Salam, A., Vuran, M. C. & Irmak, S. *Pulses in the sand: Impulse response analysis of wireless underground channel* in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications* (2016), 1–9.

37. Agilent. *33220A 20 MHz Waveform Generator* <https://web.sonoma.edu/esee/manuals/33220-90002.pdf>.

38. Clear Ballistic. *Ballistic Gelatin* <https://www.clearballistics.com/shop/10-ballistic-gelatin-fbi-block/>.

39. Othman, N. S., Jaafar, M. S., Rahman, A. A. & Sazlinayati, E. *Ultrasound propagation speed of polymer gel mimicked human soft tissue in 23 days* in *2011 International Conference on Biomedical Engineering and Technology* (2011).

40. Shepherd, C., Appleby-Thomas, G. J., Hazell, P. & Allsop, D. *The dynamic behaviour of ballistic gelatin* in *AIP conference proceedings* **1195** (2009), 1399–1402.

41. Winter, J & Shifler, D. *The material properties of gelatin gels* tech. rep. (MARVALAUD INC WESTMINSTER MD, 1975).

42. Mini-Circuits. *Low Noise Amplifiers ZFL-1000LN+.* <https://ww2.minicircuits.com/pdfs/ZFL-1000LN+.pdf>.

43. Keysight Technologies. *Keysight Infiniium Oscilloscopes - 900A series* <https://www.keysight.com/en/pcx-x205195/infiniium-90000a-series-oscilloscopes?cc=IT{\%}5C{\&}lc=ita{\%}7D>.

44. Elliott, C., Vijayakumar, V., Zink, W. & Hansen, R. National Instruments LabVIEW: A Programming Environment for Laboratory Automation and Measurement. *Journal of Laboratory Automation.* ISSN: 15402452. doi:10.1016/j.jala.2006.07.012 (2007).

45. Rappaport, T. S. *et al. Wireless communications: principles and practice* (prentice hall PTR New Jersey, 1996).

46. Treeby, B. E. & Cox, B. T. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of biomedical optics* **15,** 021314 (2010).

47. Sciacca, E., Alaimo, S., Pulvirenti, A., Latora, V., Humby, F., Ferro, A., Lewis, M. J. & Pitzalis, C. P22 Micro-RNA enriched pathway impact analysis applied to synovial RNA-seq in early rheumatoid arthritis identifies response prediction pathways. *Rheumatology* **59,** keaa111–021 (2020).

48. QIAGEN, I. *Ingenuity Pathway Analysis* <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>.

49. Alaimo, S., Giugno, R., Acunzo, M., Veneziano, D., Ferro, A. & Pulvirenti, A. Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. *Oncotarget* **7,** 54572–54582. ISSN: 19492553 (2016).

50. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H. & D'Eustachio, P. The reactome pathway knowledgebase. *Nucleic Acids Research* **48,** D498–D503. ISSN: 13624962 (2020).

51. Tarca, A. L., Draghici, S., Khatri, P., Hassan, S. S., Mittal, P., Kim, J.-s., Kim, C. J., Kusanovic, J. P., Romero, R., Hall, S. & John, R. A novel signaling pathway impact analysis. **25,** 75–82 (2009).

52. Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., Haussler, D. & Stuart, J. M. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. **26,** 237–245 (2010).

53. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28,** 27–30 (2000).

54. Hsu, S. D., Lin, F. M., Wu, W. Y., Liang, C., Huang, W. C., Chan, W. L., Tsai, W. T., Chen, G. Z., Lee, C. J., Chiu, C. M., Chien, C. H., Wu, M. C., Huang, C. Y., Tsou, A. P. & Huang, H. D. MiRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research.* ISSN: 03051048. doi:`10.1093/nar/gkq1107` (2011).

55. Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X. & Li, T. miRecords: an integrated resource for microRNA–target interactions. *Nucleic acids research* **37,** D105–D110 (2009).

56. Tong, Z., Cui, Q., Wang, J. & Zhou, Y. TransmiR v2.0: An updated transcription factor-microRNA regulation database. *Nucleic Acids Research.* ISSN: 13624962. doi:`10.1093/nar/gky1023` (2019).

57. Sahimi, M. *Applications of percolation theory* (CRC Press, 1994).

58. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43,** e47–e47 (2015).

59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15,** 550 (2014).

60. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43,** e47–e47 (2015).

61. Of Rheumatology, A. C. *ACR Convergence 2020* <`https://www.rheumatology.org/Annual-Meeting`>.

62. Sciacca, E., Surace, A., Pulvirenti, A., Alaimo, S., Ferro, A., Latora, V., Pitzalis, C. & Lewis, M. J. *TNovel Network Tool Highlights Key Features Associated with Disease Pathotypes and Response to Treatment in Early Rheumatoid Arthritis*

63. Pontarini, E., Chowdhury, F., Sciacca, E., Lucchesi, D., Rivellese, F., Goldmann, K., Lewis, M. J., Gregoriafow, S., Bowman, S., Pitzalis, C. & Bombardieri, M. *TRACTISS cohort: Rituximab prevent salivary gland inflammation in primary Sjogren's Syndrome patients*

64. Experimental Medicine Rheumatology Department, Queen Mary University London. *The PEAC cohort* <http://www.peac-mrc.mds.qmul.ac.uk/access.php>.

65. Humby, F., Lewis, M., Ramamoorthi, N., Hackney, J. A., Barnes, M. R., Bombardieri, M., Setiadi, A. F., Kelly, S., Bene, F., DiCicco, M., *et al.* Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Annals of the rheumatic diseases* **78,** 761–772 (2019).

66. Firestein, G. S. Evolving concepts of rheumatoid arthritis. *Nature* **423,** 356–361 (2003).

67. Humby, F., Lewis, M., Ramamoorthi, N., Hackney, J. A., Barnes, M. R., Bombardieri, M., Setiadi, A. F., Kelly, S., Bene, F., DiCicco, M., *et al.* Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Annals of the rheumatic diseases* **78,** 761–772 (2019).

68. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W. H., Pagès, F., Trajanoski, Z. & Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25,** 1091–1093. ISSN: 13674803 (2009).

69. Collison, J. *Extracellular matrix in the crosshairs* 2019. doi:10.1038/s41584-019-0160-7.

70. Loeser, R. F. *Integrins and chondrocyte-matrix interactions in articular cartilage* 2014. doi:10.1016/j.matbio.2014.08.007.

71. Loyher, P. L., Rodero, M. P., Combadière, C. & Boissonnas, A. Role of chemokines and chemokine receptors in cancer. *Cancer Immunology: A Translational Medicine Context, Second Edition,* 235–262 (2020).

72. Elemam, N. M., Hannawi, S. & Maghazachi, A. A. Role of Chemokines and Chemokine Receptors in Rheumatoid Arthritis. *ImmunoTargets and Therapy* **9,** 43 (2020).

73. Rajurkar, M., Dang, K., Fernandez-Barrena, M. G., Liu, X., Fernandez-Zapico, M. E., Lewis, B. C. & Mao, J. IKBKE is required during KRAS-induced pancreatic tumorigenesis. *Cancer research* **77,** 320–329 (2017).

74. Cockwell, P., Calderwood, J. W., Brooks, C. J., Chakravorty, S. J. & Savage, C. O. Chemoattraction of T cells expressing CCR5, CXCR3 and CX3CR1 by proximal tubular epithelial cell chemokines. *Nephrology Dialysis Transplantation.* ISSN: 09310509. doi:`10.1093/ndt/17.5.734` (2002).

75. Bazan, J. F., Bacon, K. B., Hardiman, G., Wang, W., Soo, K., Rossi, D., Greaves, D. R., Zlotnik, A. & Schall, T. J. A new class of membrane-bound chemokine with a CX3C motif. *Nature.* ISSN: 00280836. doi:`10.1038/385640a0` (1997).

76. Chawla, A. *Control of macrophage activation and function by PPARs* 2010. doi:`10.1161/CIRCRESAHA.110.216523`.

77. Tanaka, K., Hiraiwa, N., Hashimoto, H., Yamazaki, Y. & Kusakabe, M. Tenascin-C regulates angiogenesis in tumor through the regulation of vascular endothelial growth factor expression. *International journal of cancer* **108,** 31–40 (2004).

78. Ahnert, P. & Kirsten, H. *Association of ITGAV supports a role of angiogenesis in rheumatoid arthritis* 2007.

79. Miao, C.-g., Yang, Y.-y., He, X., Li, X.-f., Huang, C., Huang, Y., Zhang, L., Lv, X.-W., Jin, Y. & Li, J. Wnt signaling pathway in rheumatoid arthritis, with special emphasis on the different roles in synovial inflammation and bone remodeling. *Cellular signalling* **25,** 2069–2078 (2013).

80. Perlman, H., Georganas, C., Pagliari, L. J., Koch, A. E., Haines, K. & Pope, R. M. Bcl-2 Expression in Synovial Fibroblasts Is Essential for Maintaining Mitochondrial Homeostasis and Cell Viability. *The Journal of Immunology.* ISSN: 0022-1767. doi:`10.4049/jimmunol.164.10.5227` (2000).

81. Liu, H. & Pope, R. M. *The role of apoptosis in rheumatoid arthritis* 2003. doi:`10.1016/S1471-4892(03)00037-7`.

82. Smith, M. D., Weedon, H., Papangelis, V., Walker, J., Roberts-Thomson, P. J. & Ahern, M. J. Apoptosis in the rheumatoid arthritis synovial membrane: modulation by disease-modifying anti-rheumatic drug treatment. *Rheumatology (Oxford, England).* ISSN: 14620332. doi:`10.1093/rheumatology/kep467` (2010).

83. Pap, T., Müller-Ladner, U., Gay, R. E. & Gay, S. *Fibroblast biology. Role of synovial fibroblasts in the pathogenesis of rheumatoid arthritis* 2000. doi:`10.1186/ar113`.

84. Liang, Y., Xu, W. D., Peng, H., Pan, H. F. & Ye, D. Q. *SOCS signaling in autoimmune diseases: Molecular mechanisms and therapeutic implications* 2014. doi:`10.1002/eji.201344369`.

85. Greisen, S. R., Schelde, K. K., Rasmussen, T. K., Kragstrup, T. W., Stengaard-Pedersen, K., Hetland, M. L., Hørslev-Petersen, K., Junker, P., Øster-gaard, M., Deleuran, B. & Hvid, M. CXCL13 predicts disease activity in early rheumatoid arthritis and could be an indicator of the therapeutic 'window of opportunity'. *Arthritis Research and Therapy.* ISSN: 14786362. doi:`10.1186/s13075-014-0434-z` (2014).

86. Moschovakis, G. L., Bubke, A., Friedrichsen, M., Falk, C. S., Feederle, R. & Förster, R. T cell specific Cxcr5 deficiency prevents rheumatoid arthritis. *Scientific Reports.* ISSN: 20452322. doi:`10.1038/s41598-017-08935-6` (2017).

87. Cho, K., Demissie, S., Dupuis, J., Cupples, L. A., Kathiresan, S., Beck, T. J., Karasik, D. & Kiel, D. P. Polymorphisms in the endothelial nitric oxide synthase gene and bone density/ultrasound and geometry in humans. *Bone.* ISSN: 87563282. doi:`10.1016/j.bone.2007.09.051` (2008).

88. Mountz, J., Zhang, H.-G., Wang, Y, Xie, J., Liang, X, Hsu, H.-C. & Curiel, D. AKT regulates TNF-alpha-mediated apoptosis of rheumatoid arthritis synovial fibroblasts. *Arthritis research.* ISSN: 1465-9905. doi:`10.1186/ar374` (2001).

89. Zhang, X., Guo, L., Collage, R. D., Stripay, J. L., Tsung, A., Lee, J. S. & Rosengart, M. R. Calcium/calmodulin-dependent protein kinase (CaMK) Iα mediates the macrophage inflammatory response to sepsis. *Journal of Leukocyte Biology.* ISSN: 0741-5400. doi:`10.1189/jlb.0510286` (2011).

90. Brito-Zeron, P., Baldini, C., Bootsma, H., Bowman, S. J., Jonsson, R., Mariette, X., Sivils, K., Theander, E., Tzioufas, A. & Ramos-Casals, M. Sjögren syndrome. *Nature reviews Disease primers* **2,** 1–20 (2016).

91. Tzioufas, A. G., Kapsogeorgou, E. K. & Moutsopoulos, H. M. Pathogenesis of Sjögren's syndrome: what we know and what we should learn. *Journal of autoimmunity* **39,** 4–8 (2012).

92. Hansen, A., Odendahl, M., Reiter, K., Jacobi, A. M., Feist, E., Scholze, J., Burmester, G. R., Lipsky, P. E. & Dörner, T. Diminished peripheral blood memory B cells and accumulation of memory B cells in the salivary glands of patients with Sjögren's syndrome. *Arthritis & Rheumatism* **46,** 2160–2171 (2002).

93. Bowman, S. J., Everett, C. C., O'Dwyer, J. L., Emery, P., Pitzalis, C., Ng, W.-F., Pease, C. T., Price, E. J., Sutcliffe, N., Gendi, N. S., *et al.* Randomized controlled trial of rituximab and cost-effectiveness analysis in treating fatigue and oral dryness in primary Sjögren's syndrome. *Arthritis & Rheumatology* **69,** 1440–1450 (2017).

94. Fisher, B. A., Everett, C. C., Rout, J., O'Dwyer, J. L., Emery, P., Pitzalis, C., Ng, W.-F., Carr, A., Pease, C. T., Price, E. J., *et al.* Effect of rituximab on a salivary gland ultrasound score in primary Sjögren's syndrome: results of the TRACTISS randomised double-blind multicentre substudy. *Annals of the rheumatic diseases* **77,** 412–416 (2018).

95. Bootsma, H., Kroese, F. G. & Vissink, A. Rituximab in the Treatment of Sjögren's Syndrome: Is It the Right or Wrong Drug? *Arthritis & Rheumatology* **69,** 1346–1349 (2017).

96. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. & Ragg, T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology* **7,** 1–14 (2006).

97. Matsubara, T., Soh, J., Morita, M., Uwabo, T., Tomida, S., Fujiwara, T., Kanazawa, S., Toyooka, S. & Hirasawa, A. DV200 Index for Assessing RNA Integrity in Next-Generation Sequencing. *BioMed research international* **2020** (2020).

98. Storey JD Bass AJ, D. A.R. D. *qvalue: Q-value estimation for false discovery rate control. R package version 2.20.0* 2020. <http://github.com/jdstorey/qvalue>.

99. in. *The Concise Encyclopedia of Statistics* 102–102 (Springer New York, New York, NY, 2008). ISBN: 978-0-387-32833-1. doi:10.1007/978-0-387-32833-1_71. <https://doi.org/10.1007/978-0-387-32833-1_71>.

100. Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* **58,** 236–244 (1963).