AGRICULTURAL UNIVERSITY OF ATHENS
DEPARTMENT OF BIOTECHNOLOGY
GENETICS LABORATORY

Ph.D. Thesis

Genome-Wide Association Studies (GWAS) in an effort to provide insights into the complex interplay of nuclear receptor transcriptional networks and their contribution to the maintenance of homeostasis

Athanasios D. Mitsis
(Referred to as Thanasis Mitsis)

Thesis Supervisor:
Dimitrios Vlachakis, Assistant Professor, Agricultural University of Athens

Ph.D. Advising Committee:
Dimitrios Vlachakis, Assistant Professor, Agricultural University of Athens
George P. Chrousos, Professor of Emeritus, National & Kapodistrian University of Athens
Elias Eliopoulos, Professor, Agricultural University of Athens

ATHENS 2022

AGRICULTURAL UNIVERSITY OF ATHENS
DEPARTMENT OF BIOTECHNOLOGY
GENETICS LABORATORY

Ph.D. Thesis

Genome-Wide Association Studies (GWAS) in an effort to provide insights into the complex interplay of nuclear receptor transcriptional networks and their contribution to the maintenance of homeostasis

"Μελέτες συσχέτισης πλήρους γονιδιώματος (GWAS) με στόχο την εύρεση πληροφοριών για το σύστημα αλληλεπίδρασης των πυρηνικών υποδοχέων σε επίπεδο μεταγραφής και τη συμβολή του στη διατήρηση της ομοιόστασης"

Athanasios D. Mitsis

Examination Committee:
Dimitrios Vlachakis, Assistant Professor, Agricultural University of Athens (Supervisor)
George P. Chrousos, Professor of Emeritus, National & Kapodistrian University of Athens
Elias Eliopoulos, Professor, Agricultural University of Athens
Christos Yapijakis, Associate Professor, National & Kapodistrian University of Athens
Flora Bacopoulou, Associate Professor, National & Kapodistrian University of Athens
Themis Exarchos, Assistant Professor, Ionian University
George N Goulielmos, Professor, University of Crete

Genome-Wide Association Studies (GWAS) in an effort to provide insights into the complex interplay of nuclear receptor transcriptional networks and their contribution to the maintenance of homeostasis

Department of Biotechnology
Genetics Laboratory

Abstract

Transcription factors are proteins that bind short distinct DNA sequences and regulate gene expression. These proteins play a major role in organism evolution and the emergence of phenotypic variation. Nuclear receptors (NRs)are a category of ligand-dependent transcription factors that participate in essential biological processes. The glucocorticoid receptor(GR)specifically, is a nuclear receptor that modulates the stress response system and therefore plays an integral role in homeostasis maintenance. Homeostasis, as a physiological mechanism, is essential in proper organism function. GR interacts with numerous epigenetics factors, enzymes and even other nuclear receptors. Genetic or structural alterations on such transcription factors may have severe consequences on numerous biological mechanisms.

This work investigates the available genomic and structural ata on nuclear receptors in an attempt to provide novel information regarding the interplay between NR transcriptional networks. The evolutionary history of nuclear receptors based on their ligand-binding domain(LBD) structural region and the pathological conditions emerging after alterations on these regions are also studied. Since the glucocorticoid receptor has a major role in homeostasis maintenance, a data-mining and semantics pipeline was applied in order to extract important information from associated SNPs found in the current literature. Lastly, the frequencies of single nucleotide polymorphisms found in the GR interactome were used to compare two different populations, in an effort to find homeostasis-related characteristics distinct for specific individuals.

Results showcase that the ligand-binding domain of nuclear receptors displays seven evolutionary conserved signaling motifs, including the ‘LxxLL’ or ‘LLxxL’ repeating amino-acid patterns. Phylogenetic analysis revealed four monophyletic branches and hinted at new evolutionary relations among NRs’ LBD. Moreover, structural and functional comparisons on NRs’ LBD structures and their associated ligands identified two distinct canonical forms, one steroid hormone receptor-like cluster and one thyroid hormone receptor-like cluster. It was also uncovered that the estrogen

receptor alpha (ERa) was split into two distinct sub-clusters. Although these sub-clusters were very similar sequence-wise, one of them was structurally more similar to estrogen receptor beta(RMSD <2Å)than ERa. It is possible that this sub-cluster of ERa emerges due to aY537S mutation, which has been heavily associated with breast cancer.ThisY537Ssub-clusterismore similar to the estrogen receptor beta on a functional and structural level but still retains some of ERa's properties. The data-mining and semantics approach focused on single nucleotide polymorphisms found in GR and its signaling regulators highlighted the importance of this receptor in homeostasis maintenance. SNPs in intronic regions were associated with severe pathological conditions, an effect possibly achieved via the action of non-coding RNAs that interfere with gene expression. The POLR1C gene was also found to be heavily present in GR signaling regulators' literature, implying that this receptor plays an important role in rRNA synthesis. Lastly, in order to evaluate how SNPs associated with the glucocorticoid receptor may influence phenotypic variation, a genetic comparison between the Japanese and Korean populations was conducted. The comparison once again highlighted that alteration in intronic regions may lead to several pathological conditions. Additionally, the single nucleotide polymorphismrs1043618found in the HSP1A may be responsible for characteristics distinct to each population.

These results highlight the importance of nuclear receptors in numerousbiological mechanisms, including homeostasis.Additionally,uncoveredstructuralinformation maybeused for the development of novel drugs,whilenewgenetic information can helpimprove disease diagnosis.

Scientific Area: Human genetics

Keywords: transcription factors, nuclear receptors, glucocorticoid receptor, homeostasis, stress response, single nucleotide polymorphisms

Μελέτες συσχέτισης πλήρους γονιδιώματος (GWAS) με στόχο την εύρεση πληροφοριών για το σύστημα αλληλεπίδρασης των πυρηνικών υποδοχέων σε επίπεδο μεταγραφής και τη συμβολή του στη διατήρηση της ομοιόστασης

Τμήμα Βιοτεχνολογίας
Εργαστήριο Γενετικής

Περίληψη

Οι μεταγραφικοί παράγοντες αποτελούν πρωτεΐνες που δεσμεύουν συγκεκριμένες μικρές αλληλουχίες DNA και ρυθμίζουν τη γονιδιακή έκφραση. Αυτές οι πρωτεΐνες έχουν σημαντικό ρόλο στην εξέλιξη των οργανισμών και στην εμφάνιση της φαινοτυπικής ποικιλομορφίας. Οι πυρηνικοί υποδοχείς είναι μια κατηγορία μεταγραφικών παραγόντων που ρυθμίζονται από την πρόσδεση ενός συνδέτη και παίρνουν μέρος σε σημαντικές βιολογικές διεργασίες. Ο υποδοχέας των γλυκοκορτικοειδών (GR), συγκεκριμένα, είναι ένας πυρηνικός υποδοχέας που ρυθμίζει την απόκριση στο στρες και συνεπώς παίζει κύριο ρόλο στη διατήρηση της ομοιόστασης. Η ομοιόσταση, ως φυσιολογικός μηχανισμός, είναι απαραίτητη για τη σωστή λειτουργία ενός οργανισμού. Γενετικές ή δομικές αλλαγές σε τέτοιας φύσεως μεταγραφικούς παράγοντες μπορεί να έχουν δριμείς επιπτώσεις σε διάφορους βιολογικούς μηχανισμούς.

Η παρούσα εργασία ερευνά τα διαθέσιμα γενετικά και δομικά δεδομένα για τους πυρηνικούς υποδοχείς και προσπαθεί να προσφέρει καινούριες πληροφορίες που σχετίζονται με τα μεταγραφικά δίκτυα των πυρηνικών υποδοχέων και την αλληλεπίδρασή τους. Επίσης, μελετήθηκαν η εξελικτική ιστορία των πυρηνικών υποδοχέων με βάση τη δομική περιοχή δέσμευσης συνδέτη καθώς και οι παθολογικές καταστάσεις που εμφανίζονται μετά από τροποποιήσεις σε αυτές τις περιοχές. Καθώς ο υποδοχέας των γλυκοκορτικοειδών έχει σημαντικό ρόλο στη διατήρηση της ομοιόστασης, έγινε χρήση μεθόδων εξόρυξης δεδομένων και σημασιολογίας για τη λήψη σημαντικών πληροφοριών από συσχετιζόμενους μονονουκλεοτιδικούς πολυμορφισμούς που βρίσκονται στην υπάρχουσα βιβλιογραφία. Τέλος, οι συχνότητες μονονουκλεοτιδικών πολυμορφισμών που εντοπίζονται σε γονίδια που αλληλοεπιδρούν με τον υποδοχέα των γλυκοκορτικοειδών χρησιμοποιήθηκαν για τη σύγκριση δύο διαφορετικών πληθυσμών σε μια προσπάθεια εύρεσης γενετικών χαρακτηριστικών που σχετίζονται με την ομοιόσταση και εντοπίζονται σε συγκεκριμένα άτομα.

Τα αποτελέσματα παρουσιάζουν ότι η δομική περιοχή πρόσδεσης του συνδέτη των πυρηνικών υποδοχέων παρουσιάζει εφτά εξελικτικά διατηρημένες περιοχές σηματοδότησης, συμπεριλαμβανομένου του επαναλαμβανόμενου αμινοξικού μοτίβου“LxxLL”ή“LLxxL”.Μια φυλογενετική ανάλυση παρουσίασε τέσσερις μονοφυλετικούς κλάδους και προτείνει μια νέα εξελικτική σχέση μεταξύ των

περιοχών πρόσδεσης του συνδέτη των πυρηνικών υποδοχέων. Επιπροσθέτως, δομικές και λειτουργικές αναλύσεις στις περιοχές πρόσδεσης του συνδέτη των πυρηνικών υποδοχέων παρουσίασαν δύο διαφορετικές υποομάδες, μια που μοιάζει με τους υποδοχείς στεροειδών και μια που μοιάζει με τον υποδοχέα του θυρεοειδούς. Εκτός αυτού, ανακαλύφθηκε ότι ο υποδοχέα οιστρογόνων τύπου α χωρίζεται σε δύο ξεχωριστές υποομάδες. Παρότι οι υποομάδες αυτές είχαν πολλές ομοιότητες σε επίπεδο αλληλουχίας, η μια φαίνεται να μοιάζει δομικά περισσότερο με τον υποδοχέα των οιστρογόνων τύπου β παρά με τον τύπου α (RMSD <2Å).Πιθανόν η υποομάδα αυτή του υποδοχέα να εμφανίζεται λόγω της μετάλλαξηςY537Sπου έχει συσχετισθεί σε μεγάλο βαθμό με τον καρκίνο του μαστού. Αυτή η υποομάδα που δημιουργείται από την Y537S θυμίζει περισσότερο τον υποδοχέα οιστρογόνων τύπου β σε επίπεδο δομής και λειτουργίας αλλά διατηρεί ακόμα κάποια από τα χαρακτηριστικά του υποδοχέα τύπου α. Η προσέγγιση εξόρυξης δεδομένων και σημασιολογίας που εφαρμόστηκε στους μονονουκλεοτιδικούς πολυμορφισμούς που εμφανίζονται στον υποδοχέα των γλυκοκορτικοειδών και τους ρυθμιστές της σηματοδότησής του επεσήμανε το ρόλο του υποδοχέα στη διατήρηση της ομοιόστασης. Οι μονονουκλεοτιδικοί πολυμορφισμοί που εντοπίζονται σε εσωνικές περιοχές φέρεται να σχετίζονται με δριμείς παθολογικές καταστάσεις, μια ιδιότητα που μάλλον εκτελείται μέσω της δράσης μη κωδικών RNA που παρεμβάλλονται στη γονιδιακή έκφραση. Το γονίδιο POLR1C εντοπίστηκε αρκετές φορές στη βιβλιογραφία των ρυθμιστών της σηματοδότησης του υποδοχέα των γλυκοκορτικοειδών, προτείνοντας, κατά αυτόν τον τρόπο, ένα πιο σημαντικό ρόλο του υποδοχέα στη σύνθεση του rRNA. Τέλος, έλαβε χώρα μια γενετική σύγκριση μεταξύ του Γιαπωνέζικου και του Κορεάτικου πληθυσμού ώστε να διαλευκανθεί το πώς οι μονονουκλεοτιδικοί πολυμορφισμοί που σχετίζονται με τον υποδοχέα των γλυκοκορτικοειδών επηρεάζουν τη φαινοτυπική ποικιλομορφία. Η σύγκριση επισήμανε για άλλη μια φορά πώς αλλαγές σε ιντρονικές περιοχές οδηγούν σε δριμείς παθολογικές καταστάσεις. Επιπροσθέτως, ο πολυμορφισμόςrs1043618που εντοπίζεται στο γονίδιο HSP1A δύναται να ευθύνεται για χαρακτηριστικά ξεχωριστά για τον εκάστοτε πληθυσμό.

Τα αποτελέσματα αυτά επισημαίνουν τη σημαντικότητα των πυρηνικών υποδοχέων σε πολλούς βιολογικούς μηχανισμούς, συμπεριλαμβανομένης της ομοιόστασης. Επιπλέον, δομικές πληροφορίες που εντοπίστηκαν μπορούν να χρησιμοποιηθούν για την ανάπτυξη νέων φαρμάκων, ενώ γενετικές πληροφορίες που εντοπίστηκαν μπορούν να χρησιμοποιηθούν στη διάγνωση και πρόγνωση ασθενειών.

Επιστημονικό Πεδίο: Γενετική ανθρώπου

Λέξεις κλειδιά: μεταγραφικοί παράγοντες, πυρηνικοί υποδοχείς, υποδοχέας των γλυκοκορτικοειδών, ομοιόσταση, απόκριση στο στρες, μονονουκλεοτιδικοί πολυμορφισμοί

Table of Contents

Acknowledgments

Looking back at my years spent at the Agricultural University of Athens, I realize how lucky I have been to walk along with great mentors, colleagues, family, and friends. I would like to take a few lines to acknowledge the people who helped me push through a difficult time in my life.

First, I would like to thank my main supervisor Dr. Dimitrios Vlachakis, who helped me throughout the entirety of my studies, taught me how to analyze, formulate, and disseminate scientific information, and encouraged me in pursuing my ambitions.

The remaining members of my supervising committee, Dr. George P Chrousos and Dr. Elias Eliopoulos, since not many candidates get to have scientists of their caliber supervising their work and receiving their advice.

I would also like to specifically mention Dr. Louis Papageorgiou, Dr. Stefanos Leptidis, and Eleni Papakonstantinou for showing me the ropes in scientific study. These individuals took a mentoring role while I was working on my Ph.D. This text would not exist without their help, and for that, I am grateful.

I would like to thank my fellow Ph.D. candidates Katerina Pierouli and Io Diakou. We worked together on numerous projects, we got stressed, we got tired, but we managed to succeed by excellent cooperation and what I would like to call friendship.

I would like to thank my family and friends for supporting me throughout the years. Especially my parents, who, through difficult financial situations and under severe circumstances, managed to always be there for me and help me overcome some of my life's biggest obstacles.

Lastly, I would like to thank you, dear reader, for going through this text. There is a lot of work put into it, and I hope it provides you with information you will find -in some way or another- useful in the future.

Permissions

The current work was inspected with my permission by the examination committee and plagiarism software owned by the Agricultural University of Athens. Its validity and originality were confirmed.

Με την άδειά μου, η παρούσα εργασία ελέγχθηκε από την Εξεταστική Επιτροπή μέσα από λογισμικό ανίχνευσης λογοκλοπής που διαθέτει το ΓΠΑ και διασταυρώθηκε η εγκυρότητα και η πρωτοτυπία της

List of Tables

List of Figures

Introduction

Stress and Homeostasis

All living organisms need to cope with a number of challenges during their lifespan. Consequently, living organisms are programmed to maintain an inner equilibrium, both physical and psychological, despite life challenges. This inner equilibrium is called homeostasis and is essential in proper organism function (1). The internal or external unforeseen stimuli that threaten homeostasis are called stress factors or stressors, while stress is a state of threatened or perceived as threatened homeostasis. In order to cope with stressors, organisms have developed a complex system that includes both physiological and behavioral responses. This system is called the stress response system and is partly located in the central nervous system (CNS) and peripheral organs (2). The central components of the stress system, located in the hypothalamus and brainstem, include a) parvocellular neurons that release corticotropin-releasing hormone (CRH), b) paraventricular nuclei (PVN) neurons that release arginine vasopressin (AVP), c) CRH neurons of the paragigantocellular and parabrachial nuclei of the medulla and locus coeruleus (LC) and d), norepinephrine (NE) cell groups in the pons and medulla, known as the LC/NE system. The stress system's peripheral components include a) the peripheral part of the Hypothalamic-Pituitary-Adrenal (HPA) axis, b) components of the parasympathetic system, and c) the efferent sympathetic adreno-medullary (SAM) system **(Figure 1)** (3). The HPA axis is considered to be the main facilitator of the stress response system (4).

In the central nervous system, stress activates neural paths that partake in functions related to stimulation, feeding and reproduction. In peripheral organs, stress partakes in the redistribution of energy. Therefore, under stress, oxygen and various nutrients are distributed to the CNS and peripheral organs that are part of the stress response system while there's also an increase in blood pressure, pulse, breathing, gluconeogenesis and lipolysis (5).

An organism's response to stressors is influenced by both genetic and environmental factors. If the aforementioned response is not sufficient, is too severe or lasts longer than normal, there is a negative impact on essential biological processes. Some of those processes include metabolism, development, reproduction, immune system pathways and various cognitive functions. This disequilibrium in an organism leads to wear and tear on both body and brain and is termed allostatic load or overload (6). The active state adapting to stressors is called allostasis and aims to retain or reinstate homeostasis. Allostasis is regulated by the immune system, the autonomous nervous system, metabolism and hormones that are produced during stress.

Stress can be either beneficial or toxic (6). Short-term stress, also known as acute stress, generally results in adaptive changes that help maintain homeostasis, while

chronic stress has been associated with pathological conditions. Therefore, the timely cessation of the stress response is important to prevent damage associated with prolonged response (7). Beneficial stress can act as a favorable factor that extends the adaptability of an organism. Toxic stress is defined by a prolonged or permanent pathological response to stressors with a high risk of disease. Early life stress can also change neural architecture, leading to a severe response to stressors and subsequent toxic stress (8). The molecular underpinnings of stress's physiological and pathological effects are mostly associated with LC/NE system and HPA axis function.

The LC/NE, SAM and parasympathetic systems

The LC/NE system is also known as the central noradrenergic system. The locus coeruleus is a cluster of norepinephrine-producing neurons that are located in the upper dorsolateral pontine tegmentum (9). These neurons are defined by extensively branched axons which project all through the neuraxis. They are the sole source of NE to the hippocampus, neocortex, cerebellum and most of the thalamus. The SAM system function, also known as the peripheral noradrenergic system, is largely controlled by the LC/NE system and includes the NE neurons of the sympathetic system and the adrenal medulla (10). Adrenal medulla stimulation by the LC/NE system leads to catecholamines secretion (CEs), specifically epinephrine (E) and norepinephrine (NE) (11). The sympathetic system is essential in the 'fight or flight' response to stress by secreting the aforementioned catecholamines, epinephrine (also known as adrenaline) and norepinephrine (also known as noradrenaline). The parasympathetic system assists or antagonizes sympathetic functions by withdrawing or increasing its activity respectively (2). Therefore, the sympathetic and parasympathetic system interact in order to regulate the "fight or flight" response. The co-ordination of the "fight or flight" response is an integral part of stress response, which, along with HPA axis function, has been the object the object of many studies.

The HPA axis

A perceived stressor induces a chain of events in the brain that signals the paraventricular nucleus of the hypothalamus. The hypothalamus secretes CRH and AVP into the hypophysial portal system, which end up at the anterior lobe of the pituitary gland. There, these hormones will stimulate pro-opiomelanocortin (POMC) cells, which in turn will release the adrenocorticotropic hormone, also known as corticotropin (ACTH). Finally, ACTH is released into the bloodstream and acts on the adrenal glands' cortex, triggering the production of corticosteroids, more specifically glucocorticoids (cortisol in humans and corticosterone in rats) (12, 13).

The CRH released into the hypophysial portal system is the main regulator of pituitary ACTH secretion. CRH can lead sufficiently to the secretion of ACTH on its own, while AVP influences ACTH secretion on a very low level. CRH and AVP act

synergistically and they appear to interact at a molecular level in the hypothalamus, triggering each other's secretion (13). In the absence of stress, both CRH and AVP are secreted into the hypophysial portal system in a circadian and pulsatile fashion (14). The amplitude of CRH and AVP pulses increases in the early morning hours, leading to the increase of both the altitude and frequency of ACTH and cortisol (CORT) secretory bursts in system circulation (13). Acute stress increases the amplitude and synchronicity of CRH and AVP pulses from the paraventricular nuclei into the hypophysial portal system. Depending on stress type, factors such as angiotensin II, cytokines and lipid mediators of inflammation may be secreted and affect the HPA axis, mostly enhancing its activity (13).

The pituitary gland secretes ACTH into system circulation, which in its turn targets the cortex of the gland. The aforementioned hormone is the main regulator of glucocorticoid and androgen gland secretion by the zonae fasciculata and reticularis, respectively. ACTH also has a role in aldosterone secretion by the zona glomerulosa. Apart from ACTH, various hormones and cytokines originating from the adrenal medulla or systemic circulation appear to partake in the regulation of cortisol secretion (13).

Glucocorticoids – as the final products of the HPA axis – play an important role in the stress response system. Glucocorticoids have the ability to inhibit HPA axis function in both a genomic and non-genomic manner through a negative feedback loop (15). This ability is mediated through the binding of glucocorticoids to their respective receptors, subsequently repressing corticotropin releasing hormone, type 1 CRH receptor and POMC genes (15). These hormones can also inhibit the LC/NE component of the stress response system, the beta-endorphin system, and stimulate the mesocorticolimbic dopaminergic system (16). Additionally, they have a major effect on the reproductive system since they can inhibit pituitary gonadotropins, growth hormone (GH), and (thyroid-stimulating hormone) TSH secretion (16).

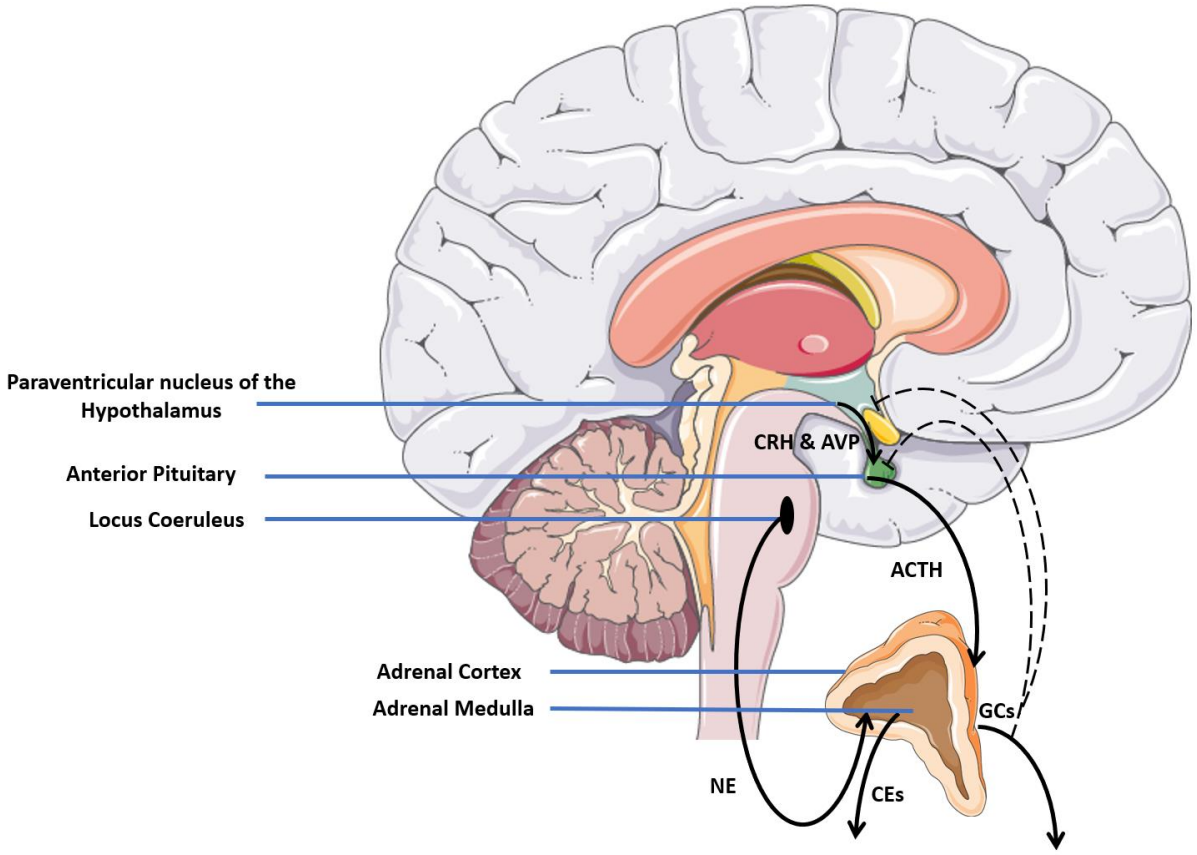

Figure 1. A schematic representation of the stress response system. HPA axis: Arginine-vasopressin (AVP) and corticotropin-releasing hormone (CRH) are secreted by the paraventricular nucleus of the hypothalamus (PVN) and stimulate the secretion of adrenocorticotropic hormone (ACTH) from the anterior pituitary. ACTH, in turn, stimulates the release of glucocorticoids (GCs) from the adrenal cortex. Glucocorticoids can inhibit HPA axis function by suppressing CRH and ACTH release. LC/NE and SAM system: The locus coeruleus (LC) consists of norepinephrine (NE) producing neurons whose axons branch throughout the neuraxis and regulate the somatic adrenomedullary system (SAM), where adrenal medulla stimulation leads to catecholamines (CEs) secretion.

Nuclear receptors and stress

Glucocorticoids act by binding to two types of intracellular receptors, mainly the glucocorticoid receptor (GR) and, on a lesser note, the mineralocorticoid receptor (MR) (17). GR is almost exclusively activated by glucocorticoids, while MR can bind GCs and the mineralocorticoid aldosterone with similar high affinity (18). MR and GR expression is similar in the gastrointestinal system, where they showcase high expression and the endocrine, metabolic, reproductive and cardiovascular systems, where they exhibit moderate expression. MR displays a higher expression than GR in the CNS and skeletal system, whereas GR displays a higher expression than MR in the immune system (19). Regarding the brain, an essential component of the stress response system, MR, as a high-affinity receptor, is occupied at basal glucocorticoids levels in the brain, while GR is activated at the circadian peak of glucocorticoid secretion and during stress (20). These observations serve to show that the main mediator of glucocorticoid action is the glucocorticoid receptor.

Both mineralocorticoid receptors and glucocorticoid receptors are part of the steroid hormone receptors class and belong to the nuclear receptors (NRs) superfamily of proteins (21). NRs are some of the most biologically important transcription factors, and they regulate the expression of a wide variety of genes. These receptors' actions are regulated by binding to small molecules and ligands **(Table 1)** (22). A wide range of lipophilic ligands can bind nuclear receptors, such as the aforementioned steroids, plus retinoids, dietary lipids, and thyroid hormones (23). This superfamily contains over 500 protein members and is divided into four classes based on characteristics such as DNA binding motifs and specificity, ligand binding, and dimerization. Those classes are steroid receptors, which comprise class I, RXR heterodimers, which comprise class II, homodimeric orphan receptors, which comprise class III and monomeric orphan receptors, which comprise class IV (24).

Table 1. A list of human nuclear receptors and their corresponding ligands

Nuclear Receptor	**Gene Name**	**Ligands**
Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX1)	NR0B1	Orphan
Short heterodimeric partner (SHP)	NR0B2	Orphan
Thyroid hormone receptor alpha (THRa)	NR1A1	Thyroid hormones
Thyroid hormone receptor beta (THRb)	NR1A2	Thyroid hormones
Retinoic acid receptor alpha (RARa)	NR1B1	Retinoic acids
Retinoic acid receptor beta (RARb)	NR1B2	Retinoic acids
Retinoic acid receptor gamma (RARg)	NR1B3	Retinoic acids
Peroxisome proliferator-activated receptor alpha (PPARa)	NR1C1	Fatty acids Eicosanoids
Peroxisome proliferator-activated receptor beta/delta (PPARb/d)	NR1C2	Fatty acids Prostaglandins
Peroxisome proliferator-activated receptor gamma (PPARg)	NR1C3	Fatty acids Prostaglandins Eicosanoids
Rev-Erb alpha (Rev-Erba)	NR1D1	Heme
Rev-Erb beta (Rev-Erbb)	NR1D2	Heme
RAR-related orphan receptor alpha (RORa)	NR1F1	Sterols
RAR-related orphan	NR1F2	Sterols

receptor beta (RORb)		
RAR-related orphan receptor gamma (RORg)	NR1F3	Sterols
Liver X receptor beta (LXRb)	NR1H2	Oxysterols
Liver X receptor alpha (LXRa)	NR1H3	Oxysterols
Farnesoid X receptor (FXR)	NR1H4	Bile Acids
Farnesoid X receptor beta (FXRb)	NR1H5P	Orphan
Vitamin D receptor (VDR)	NR1I1	Calcitriol (active form of vitamin D)
Pregnane X receptor (PXR)	NR1I2	Endobiotics and xenobiotics
Constitutive androstane receptor (CAR)	NR1I3	Xenobiotics
Hepatocyte nuclear factor 4 alpha (HNF4a)	NR2A1	Fatty acids
Hepatocyte nuclear factor 4 gamma (HNF4g)	NR2A2	Fatty acids
Retinoid X receptor alpha (RXRa)	NR2B1	9-Cis retinoic acid
Retinoid X receptor beta (RXRb)	NR2B2	9-Cis retinoic acid
Retinoid X receptor gamma (RXRg)	NR2B3	9-Cis retinoic acid
Testicular receptor 2 (TR2)	NR2C1	Orphan
Testicular Receptor 4 (TR4)	NR2C2	Orphan
Tailless homolog orphan receptor (TLX)	NR2E1	Orphan
Photoreceptor-cell-specific nuclear receptor (PNR)	NR2E3	Orphan
Chicken ovalbumin upstream promoter-transcription factor alpha (COUP-TFa)	NR2F1	Orphan
Chicken ovalbumin upstream promoter-transcription factor beta (COUP-TFb)	NR2F2	Orphan
V-erbA-related protein 2 (EAR2)	NR2F6	Orphan
Estrogen receptor alpha (ERa)	NR3A1	Estrogens

Estrogen receptor beta (ERb)	NR3A2	Estrogens
Estrogen-related receptor alpha (ERRa)	NR3B1	Orphan
Estrogen-related receptor beta (ERRb)	NR3B2	Orphan
Estrogen-related receptor gamma (ERRg)	NR3B3	Orphan
Glucocorticoid receptor (GR)	NR3C1	Glucocorticoids
Mineralocorticoid receptor (MR)	NR3C2	Mineralocorticoids and glucocorticoids
Progesterone receptor (PR)	NR3C3	Progesterone
Androgen receptor (AR)	NR3C4	Androgens
Nerve growth Factor 1B (NGF1B)	NR4A1	Orphan
Nurr-related Factor 1 (NURR1)	NR4A2	Unsaturated fatty acids
Neuron-derived orphan Receptor 1 (NOR-1)	NR4A3	Orphan
Steroidogenic Factor 1 (SF1)	NR5A1	Phospholipids
Liver receptor Homolog-1 (LRH-1)	NR5A2	Phospholipids
Germ cell nuclear factor (GCNF)	NR6A1	Orphan

Despite their differences in size and activating ligands, almost all NRs share a common modular structure **(Figure 2)**. Specifically, the regions which make up a nuclear receptor are the N-terminal domain (NTD), the DNA-binding domain (DBD), the hinge region (HR), the ligand-binding domain (LBD) and a C-terminal domain (CTD). The N-terminal domain is highly variable among nuclear receptors and contains the ligand-independent activation function-1 (AF-1) region (25). The AF-1 region interacts with various coregulators in a promoter- and cell-specific manner to modulate receptor function. The N-terminal domain is also the target for several post-translational modifications that alter receptor action, like phosphorylation, SUMOylation and acetylation (26). The DNA-binding domain is the most conserved region among nuclear receptors' domains. DBD allows NRs to bind with high specificity a group of DNA sequences, called hormone-response elements. The DNA-binding domain consists of two zinc finger modules of distinct conformation. The zinc atom is bound to four cysteine residues. The zinc fingers fold towards each other to create a single globular structure containing two helices, which are located at the C-terminal end of the zinc fingers. These helices are oriented perpendicular to each other and form a hydrophobic core. Specifically, the first helix is located in the major groove of the DNA helix with several amino acids making base-specific contacts and lays the foundation for the sequence-specific recognition of the

hexameric half-site of the hormone response elements. A loop of five amino acids, located between the first two cysteine residues in the second zinc finger, forms the core part of a homodimerization surface, with the helix at the end of the second zinc finger being responsible for several less non-specific interactions with the DNA backbone (27). The hinge region is a short and flexible region that connects the DBD and the LBD. HR displays the least sequence conservation among NRs and is a site for regulatory post-translational modifications. Additionally, this region may also contain a nuclear localization signal (26). The LBD is the second most conserved region in NRs and is responsible for binding lipophilic ligands and activating or repressing the transcriptional activities of a nuclear receptor. The LBD may be considered a molecular switch that interprets ligand structure and depicts that information as conformational changes that convert the receptor into a transcriptional activator or repressor (25). The LBD contains a common fold which has been described as a triple-helical sandwich and most often consists of twelve helices and one beta-sheet that is normally comprised of two short strands (28). The ligand-binding pocket of the receptor resides in the interior of the helical sandwich and is formed by a subset of the surrounding helices. The LBD also contains an activation function-2 (AF-2) which can recruit receptor cofactors. The structural interface for this function resides in a hydrophobic groove formed by several helices of the ligand-binding domain, including helix 12, with cofactors binding to the specific site through an LxxLL motif and via hydrophobic interactions (29). Helix-12 plays an important role in LBD function. In the absence of a ligand, helix-12 adopts a conformation that favors the interactions of the LBD with corepressor proteins, while ligand binding leads to conformational changes that expose interaction sites for coactivators, the recruitment of these coactivators and the subsequent initiation of the signal cascade that ends at target gene transcription (30). Both the DBD and the LBD contain dimerization interfaces and thus participate in receptor homodimerization and heterodimerization (25). The C-terminal domain resides at the extreme C-terminus of an NR while its high sequence variability has made the elucidation of its function difficult (25).

Figure 2. A schematic representation of the common nuclear receptor structure. NTD : N-terminal Domain; AF-1 : Activation function-1; DBD : DNA-binding Domain; HR : Hinge Region; LBD : Ligand-binding domain; AF-2: Activation function-2; CTD : C-terminal domain

Apart from GR and MR, several other NRs have also been associated with the stress response system. It is known that the stress response may interfere with reproductive physiology and associated behaviors (31). This association between the stress response system and reproduction is due to the ability of the HPA axis to inhibit the hypothalamic-pituitary-gonadal (HPG) axis and vice versa. The androgen receptor plays an important role in this systems interaction (32). Estrogen receptors also have been shown to influence stress response. Specifically, estrogen receptor alpha and estrogen receptor beta work in opposition through different neuron populations in or near the PVN to regulate neuroendocrine response. A prime

example is the ability of ERb to modulate the action of neuropeptide promoters such as the CRH one (33). Apart from steroid receptors, other nuclear receptors like PPARg seem to partake in the stress response system. Particularly, PPARg signaling is thought to mitigate brain activation of the HPA axis and sympathetic nervous system during acute stress response (34). In general, there exists a complex interplay between several nuclear receptors with mentioned interactions contribute to their ability to orchestrate numerous physiological processes, including the stress response system.

Genetics, Epigenetics, and Stress

Specific interactions between environmental factors and genetic variants may lead to an altered stress response, both in a physiological and a behavioral setting (35). Therefore, genetics play an important role in stress biology since variations in genes involved in the sympathetic system or in the HPA axis may influence susceptibility to stress and lead to neuropsychiatric disorders (36). Polymorphisms within genes coding for HPA axis proteins such as the type 1 CRH receptor (CRHR1) and GR seem to lead to individual differences in stress responsivity and influence the effects of environmental stress on brain structure (37). Such an example is the genetic variant rs12938031 found in the CRHR1, which has been implicated in promoting psychopathology in the context of stress (38). Specifically, a study by Bogdan et al. found that rs12938031 interacted with stress to alter reward learning, both in a behavioral and neural way and promote stress-induced deficits in reward learning (38). Regarding the glucocorticoid receptor, individuals carrying the 22/23EK or the 9-beta variant that have been exposed to childhood adversities seem to display an increased risk of developing depression later on in their lives (39). Additionally, polymorphisms found on the FKBP5 gene, a GR cofactor and thus a modulator of the stress response system, in conjunction with early life trauma seem to be associated with neuroticism (40).

Epigenetics, i.e., the study of mechanisms and molecules that have the ability to perpetuate alternative gene activity states within the context of the same DNA sequence, also has a major influence on stress biology research (41). Particularly, epigenetic mechanisms allow long-term phenotypic changes to arise from environmental influence (42). Several molecular mechanisms can act on an epigenetic level, including DNA methylation, histone modifications, non-coding RNA (ncRNA) regulation of gene expression [4] and even protein-protein interactions (43). DNA methylation, which mainly involves the methylation of cytosine found on the DNA, is one of the most studied epigenetic mechanisms. This methylation occurs mostly at cytosines, followed by guanine residues. Generally, the presence of methylated cytosine leads to the recruitment of repressor complexes that regulate gene transcription by reshaping chromatin and inhibiting transcription factors' access to gene promoters (44). A family of enzymes named DNA methyltransferases (DNMTs) are the main modulators of DNA methylation (45). Several stressors have been shown to influence the DNA methylation patterns of genes participating in the HPA axis. Specifically, rodents exposed to maternal deprivation or chronic social

defeat stress early on in life displayed reduced DNA methylation at the transcription enhancer region of the AVP gene and the promoter region of the CRH gene (46). This decrease in methylation led to an increased expression of AVP and CRH in the hypothalamus culminating in elevated corticosterone levels and HPA axis hyperactivity. This disruption of proper HPA axis function has been associated with various neuropsychiatric disorders (46).

Metabolism and Stress

The stress response system is characterized by the increased secretion of epinephrine by the adrenal medulla, which is the final product of the SAM system and glucocorticoids from the adrenal cortex, which is the final product of the HPA axis. These stimuli may reduce insulin sensitivity, while glucocorticoids, specifically, have been associated with metabolic disturbances (47). Particularly, physiological levels of GCs are a prerequisite for proper metabolic control, while excessive GCs have been associated with various pandemic metabolic diseases like type II diabetes and obesity (48). The above indicates that stress has a major role in metabolism. Some important participators in metabolic processes that seem to be heavily associated with stress are the cytochrome P450 enzymes (CYPs) (49). CYPs are membrane-bound hemoproteins that have an essential role in homeostasis, cellular metabolism, drug metabolism along with the detoxification of xenobiotics (50). Cytochrome P450 enzymes both partake in the production of glucocorticoids and are influenced by GR function. Glucocorticoids production takes place at the adrenal zona fasciculata. Specifically, CYP11A1 catalyzes the side-chain cleavage of cholesterol to produce pregnenolone. Pregnenolone is later converted to 17α-hydroxyprogesterone (17OHP) by 3-beta-hydroxysteroid dehydrogenase (HSD3B2) and CYP17A1 17α-hydroxylase activity. Subsequently, CYP21A2 catalyzes the conversion of 17OHP to 11-deoxycortisol, with the final step being the conversion of 11-deoxycortisol to cortisol by CYP11B1, which is located in the mitochondria of zona fasciculata cells (51). Additionally, the glucocorticoid receptor seems to interfere with CYPs' function. This interference may alter the pharmacokinetic attributes of several drugs that are catalyzed by the aforementioned enzymes (52). This interaction between CYPs and GR may explain how stress can alter an organism's drug response (53).

Stress and Reproduction

Both male and female reproductive systems are inhibited at all levels by several modulators of the stress response system. Specifically, CRH suppresses gonadotropin-releasing hormone (GnRH) neurons directly and indirectly by promoting β-endorphin secretion by the arcuate POMC neurons, Additionally, GCs exhibit inhibitory effects on GnRH neurons, pituitary gonadotrophs as well as directly on the gonads, while they simultaneously render target tissues resistant to sex steroids (54). These actions of stress modulators may lead to hypothalamic amenorrhea of stress, a condition observed in various disorders such as depression and eating disorders (55).

Stress and Growth

Growth is also severely influenced by the stress response system. During stress response onset or after acute glucocorticoids administration, there's a transient elevation of growth hormone (GH) plasma levels. On the other hand, prolonged HPA axis activity suppresses GH secretion and inhibits growth factors' effect on target tissue via GC signaling. This negative effect of the prolonged stress response has been implicated in conditions such as psychosocial dwarfism, where severe childhood growth arrest or delayed puberty is observed due to emotional deprivation or harassment (54).

The Immune System and Stress

The stress response system is one of nature's essential survival mechanisms, with short-term stress mobilizing distinct cell types in the bloodstream in order to prepare the immune system for challenges, like wounding or infection, that may be inflicted by stressors, such as a predator or a medical procedure (56). Specifically, the stress response system has an important role in inflammation, a necessary short-term response for eliminating pathogens and initiating the healing process. Acute stress increases pro-inflammatory cytokines blood levels. Chronic stress is also associated with higher levels of pro-inflammatory cytokines, but the health consequences may be different. Chronic stress may lead to chronic or systemic inflammation, which can increase the risk for chronic diseases. Additionally, chronic stress may activate latent viruses, with their frequent activation causing wear-and-tear on the immune system (57).

Glucocorticoids-both endogenous and pharmacological- as the final product of the HPA axis partake in several immune and inflammatory processes. Specifically, low levels of endogenous glucocorticoids stimulate the immune response system by upregulating immune system modulators such as cytokine receptors, pattern recognition receptors (PRRs) and complement factors, therefore allowing an organism to rapidly respond to danger signals. On the other hand, high levels of glucocorticoids suppress the aforementioned immune system modulators, thus preventing excessive or prolonged immune responses (58). This anti-inflammatory action of glucocorticoids has been the basis for the frequent use of synthetic glucocorticoids as anti-inflammatory drugs against inflammatory or autoimmune diseases such as Crohn's disease, chronic obstructive pulmonary disease (COPD), rheumatoid arthritis (RA) and multiple sclerosis (MS) (59).

Stress and Brain Function

It has been briefly mentioned that the brain holds an important role in the stress response system. The neural circuity found in the brain dictates which stimuli are threatening and therefore stressful to an organism. The brain regulates physiological and psychological responses to stressful stimuli through its interaction with

metabolic, immune, and cardiovascular systems (60). A brain's reaction to stimuli, though, differs among individuals. These differences emerge due to both epigenetic factors, such as an individual's life experiences along with genetic factors, such as an individual's genetic makeup (61). Thus, the brain may be considered the quintessential organ when it comes to how an individual will respond to stress.

On the other hand, stress itself can affect brain function. Chronic stress may lead to structural changes in the brain, such as alterations in neurogenesis, a decrease in neuron numbers, in addition to disruption in memory and cognitive functions (62). Additionally, elevated glucocorticoid levels have been associated with neurodegenerative processes and thus may have a potential role in neurodegenerative diseases such as Parkinson's disease (PD) and Alzheimer's disease (AD) (63). Regarding Alzheimer's specifically, it has been suggested that long-term exposure to stress or stress-related disorders promotes hallmarks of AD-like cognitive impairment, neuroinflammation processes, amyloid-beta accumulation, Tau hyperphosphorylation and excitotoxicity (64). Lastly, chronic stress may play a role in the emergence of mood disorders like depression, where high levels of glucocorticoids contribute to neuronal atrophy in specific brain areas resulting in a brain phenotype similar to that of depressed patients (65).

Additionally, stress has a direct effect on neuronal structure and brain architecture. The main brain regions influenced by stress are the hippocampus, the amygdala, and the prefrontal cortex (PFC) (66). The hippocampus participates in memory, navigation and cognition (67). Stress leads to numerous structural changes in the hippocampus, such as neuronal morphology alterations, suppression of neuronal proliferation, and overall hippocampal volume reduction. These changes lead to the impairment of memory tasks dependent on proper hippocampal function (68). The amygdala is a limbic system with a crucial role in stress response where it detects a stimulus perceived as stressful by the organism and initiates adaptive responses, while amygdala-dependent cognition is promoted in stressful conditions. Animal models showcase that prolonged stress increases several measures of amygdala structure in rodents, such as increased arborization and dendritic length in the basolateral complex of the amygdala (69, 70). On the contrary, human studies display conflicting results (71). Nevertheless, most studies do highlight a change in amygdala morphology (72). The prefrontal cortex is a brain region with a critical role in self-regulatory behaviors, working memory, and executive function (66). Prolonged stress exposure leads to impairment in cognitive functions regulated by the PFC, while animal models have shown that such exposure promotes loss of dendritic material (73).

Stress and Neuropsychiatric Disorders

It has been stated that abnormalities in HPA function have been associated with neuropsychiatric disorders such as unipolar and bipolar depression, post-traumatic stress disorder (PTSD) and schizophrenia (74). Major depressive disorder (MDD), specifically, is a highly heterogeneous disease characterized, among others, by

depressed mood, anhedonia and anergia and may be considered a prime example of stress-related disease (75). HPA axis malfunction is present in many cases of MDD and is related to reduced feedback inhibition by glucocorticoids, resulting in axis hyperactivity with depressed patients showcasing elevated cortisol levels in plasma, saliva and urine (74, 76). These depressed patients display an increased HPA response to psychosocial stressors with deleterious effects on both cognition and physiology, while GR antagonists such as mifepristone showcase some efficacy in treating depressive symptoms (76).

Stress and specific Chronic Diseases

Chronic diseases place great health, societal, and financial burden globally. In the US specifically, most deaths are caused by distinct chronic diseases such as cancer, cardiovascular diseases, chronic obstructive pulmonary disease, and type 2 diabetes mellitus (77). As has been briefly mentioned above, stress has been associated with numerous chronic diseases and may play a pivotal part in the increasing prevalence of chronic diseases.

The relationship between stress and cancer has been the focus of numerous studies, though results seem conflicting. Glucocorticoids' effect on tumor progression appears to heavily rely on the cells targeted. In lymphocytic malignancies, dexamethasone, a synthetic glucocorticoid, is used to promote apoptotic cell death, while in epithelial cells tumors, GCs mostly display the opposite effect (78). Moreover, studies of GCs' effect on various stages of cancer paint an inconsistent picture. Glucocorticoids appear to suppress cancer metastasis. In vitro studies have shown that GCs suppress cell migration and invasion via downregulation of Ras homolog family member A (RhoA), matrix metalloproteinase 2 (MMP2), matrix metalloproteinase 9 (MMP9), and interleukin-6 (IL-6), or induction of E-cadherin (79). In contrast, it has been speculated that changes on corticosteroids levels caused by chronic stress may play a role in carcinogenesis. High glucocorticoid levels potentially enhance murine double minute 2 (MDM2) activity through serum and glucocorticoid-regulated kinase 1 (SGK1) induction (80). MDM2 is a negative regulator which plays a pivotal role in p53 inhibition (80). P53 is a protein encoded by the TP53 gene, which is a crucial tumor suppressor gene. Specifically, p53 is a main regulator of genome stability controlling cell cycle and DNA repair, thus playing a major role in tumorigenesis prevention(81). Consequently, inhibition of p53 plays an essential role in tumorigenesis, a process which high glucocorticoids levels may promote.

Stress has been identified as an important risk factor for cardiovascular diseases' emergence and progress. Exposure to chronic stress has been associated with atherosclerosis and subclinical coronary heart disease, while stressful events may hinder recovery in patients who have survived a stroke or acute coronary syndrome (82). This effect of stress is mediated both by behavioral responses to stressful events, such as consumption of energy-dense comfort food, that is interlinked with obesity and biological responses, such as the increased SAM activity, which leads to

increased heart rate, diastolic and systolic blood pressure, plus alterations on myocardial contractility and vasodilation/vasoconstriction (83-85).

Serum cortisol levels and adrenal gland size have been heavily associated with the severity of chronic obstructive pulmonary disease, an observation which implies a role for HPA axis function and subsequently the stress response system, in disease pathology (86). COPD seems to influence HPA function and psychological distress is a major symptom of this disease, but the association between stress and COPD may be bilateral since it has been suggested distress itself may be a predictor for disease emergence in women (86-88).

Physiological stress response caused by chronic exposure to stressors is thought to increase type 2 diabetes mellitus (T2DM) risk (89). Moreover, several research studies have showcased that stressors such as family losses and workplace stress may play an important role in triggering the onset of T2DM (90). Stress mobilizes several processes which have been implicated in T2DM, such as glucose and lipids release into the circulation, inflammatory cytokine expression, along with blood pressure increase. Chronic stress can lead to glucose metabolism irregularities and neuroendocrine function dysregulation, plus low-grade chronic inflammation. These pathological traits emerging from chronic stress can have major effects on T2DM since they promote hepatic gluconeogenesis activation, downregulate lipid uptake and insulin-stimulated glucose transport thus potentially having a direct effect on insulin sensitivity (91).

The glucocorticoid receptor as the main mediator of the stress response system

The glucocorticoid receptor is produced by a single gene called NR3C1, which is located on chromosome 5q31‑32 in humans. This gene is subject to alternative splicing, a process that leads to functionally distinct GR subtypes. Human NR3C1 contains nine exons, with the predominant isoforms being hGRa and hGRb. hGRa and hGRb are identical through amino acid 727, but later deviate. hGRα displays 50 additional amino acids, while hGRβ showcases 15 additional non‑homologous amino acids. hGRα is the most well-studied isoform, with the less-studied hGRβ exhibitng a dominant-negative effect on hGRα (92). The glucocorticoid receptor's mechanism of action is similar to that of other steroid receptors. In the absence of glucocorticoids, GR is located in the cytoplasm, where it is bound to several chaperone proteins that render it inactive (93). GR is first bound by heat shock protein 40 kDa (Hsp40), heat shock protein kDa (Hsp70) and the Hsp70-Hsp90 organizing protein (Hop), while at later stages is bound by heat shock protein 90 (Hsp90), FK506-binding proteins (FKBPs) and prostaglandin E synthase 3 (PTGES3/p23) (94). Specifically, after receptor translation, Hsp70 binds the unfolded receptor in the cytosol, a process accelerated by Hsp40 binding. This action facilitates GR's folding. A cofactor called BAG family molecular chaperone regulator 1 (BAG-1) may impair mentioned receptor folding, either directly or by aiding in the degradation of the unstable folded GR complex with Hsp70 and Hsp40 (95). The

Hsp40/Hsp70-GR complex is then recruited by Hop, in an ATP-dependent manner, to interact with Hsp90. Hop, Hsp40, and Hsp70 are dislodged from the Hsp90-GR complex after another ATP- dependent event and subsequent interaction of the Hsp90-GR complex with cochaperones like FKBP51 and p23 give rise to a complex conformation with a high affinity for corticosteroids (15, 96). Ligand binding leads to conformational alterations in the LBD that change the proteins which make up the heterocomplex, a prime example being the replacement of FKBP51 by FKBP52, leading -mostly- to GR dimerization and nuclear translocation, where the receptor may act as a transcriptional regulator (15, 97, 98). GR nuclear import is a quick and active process that relies on the glucocorticoid receptor's association with Hsp90, FKBP52 and importin-a. The GR complex is transported into the nucleus by dynein along the cytoskeleton and through the nuclear pore complex (NPC) (98). Once in the nucleus, the activated GR can either enable or repress gene transcription **(Figure 3)**. Especially, transactivation can be achieved directly through GR homodimer binding
to a glucocorticoid response element (GRE) found in gene promoter regions, or
indirectly, where GR acts as a monomer and co-operates with other transcription factors to induce transcription (99, 100). Transrepression can also be either direct via GR homodimer or -preferably- monomer binding to a negative glucocorticoid response element (nGRE); or indirect via GR monomer binding to a proinflammatory transcription factor, like (nuclear factor kappa B) NF-κB (99-101). It is important to highlight the glucocorticoid receptor's ability to rely on protein-protein interactions with other transcription regulators to exert a large part of its actions since transgenic mice carrying a mutant GR weakened in its ability to bind DNA but not to other proteins, are viable (48). GR remains bound to DNA for a specific time period which may be affected by the bound ligand. This influence may be due to differences in ligand-induced conformational changes (102). After ligand disengagement, GR dissociates from DNA and is either degraded by the proteasome or exported from the nucleus, which is an inactive process, most likely occurring through passive diffusion (98). This system enables the cell to rapidly respond to environmental changes and exercise its effects via the intricate networks established around GR activity.

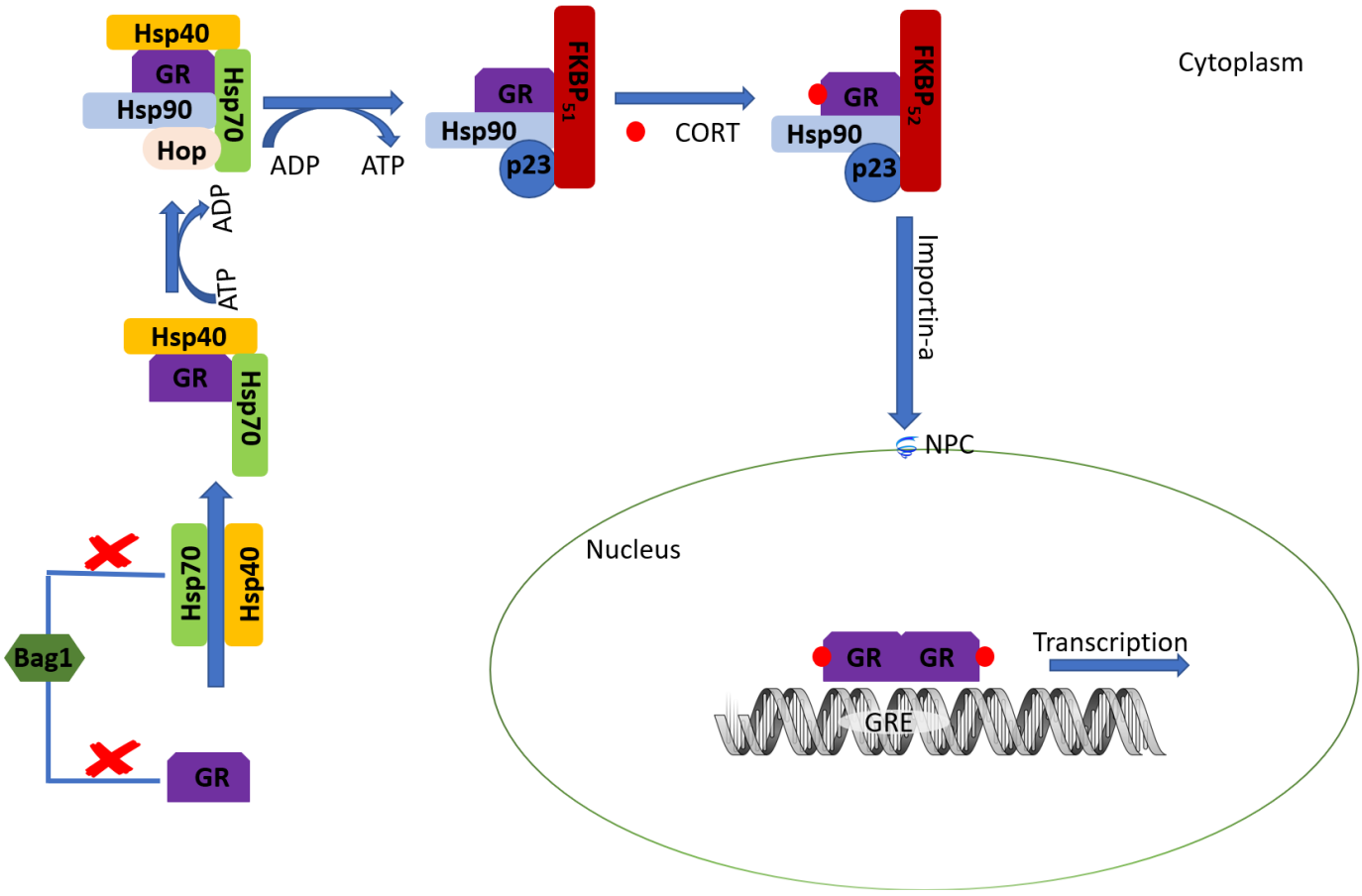

Figure 3. A schematic representation of glucocorticoid signaling resulting in GR homodimerization and transcription initiation (GR: glucocorticoid receptor; Hsp70: Heat-shock protein70; Hsp40: Heat shock protein 40; BAG1: BAG family molecular chaperone regulator 1; Hop: Hsp70-Hsp90 organizing protein; p23: Prostaglandin E synthase 3 protein; FKBP51: FK506-binding protein 51 CORT: cortisol; FKBP52: FK506-binding protein 51; NPC: nuclear pore complex; GRE: Glucocorticoid response elements). Hsp70 binds the unfolded receptor in the cytosol, a process accelerated by Hsp40 binding, and leads to GR's folding. BAG-1 is a cofactor that may directly impair receptor folding, while it may also aid in the degradation of the unstable folded GR complex with Hsp70 and Hsp40. The interaction of GR The Hsp40/Hsp70-GR complex is then recruited by Hop, in an ATP-dependent manner, to interact with Hsp90. Hop, Hsp40, and Hsp70 are dislodged from the Hsp90-GR complex upon Hsp90 binding ATP, and subsequent interaction of the Hsp90-GR complex with FKBP51 and p23 give rise to a complex conformation with a high affinity for corticosteroids. Ligand binding leads to the replacement of FKBP51 by FKBP52, leading mainly to GR dimerization and nuclear translocation through the nuclear pore complex (NPC) with the help of importin-a. Finally, the GR homodimer binds to glucocorticoid response elements to promote gene transcription.

Big Data as the basis of modern research

Current genomics and post-genomics technologies have allowed for the gathering of raw data regarding biochemical and regulatory processes in living organisms. Moreover, the use of electronic health records (EHRs) has added even more data, such as an individual patient's diagnoses, prescriptions and lab test results (103).

This accumulation of large datasets has set in motion the age of Big Data in medical and biological research. Big data refers to information assets defined by high volume, velocity and variety that demand specific technology for its proper storage, management, and analysis (104). Analyzing such data can provide new approaches for personalizing prognosis, diagnosis and therapeutics (105). One of the technological approaches used to analyze big data is through the use of artificial intelligence (AI) (106). Specifically, artificial intelligence could be used to analyze biological and medical big data and then interpret pathological results with an end goal of providing a highly accurate diagnosis (107). Furthermore, analysis of big data can build predictive models that allow an accurate disease risk and reoccurrence assessment and improve prognosis estimations (108). Regarding therapeutics, identification of factors driving a disease may provide novel pharmacological targets or help optimize the therapeutic approach to a patient based on their individual characteristics (104).

A provider of such large amounts of data is genome-wide association studies. Genome-wide association studies (GWAS) are well-powered systematic surveys that study the association between sites of common genome sequence variation and complex traits or diseases on a genome-wide scale (109). Since the first human genome was fully sequenced in 2003, almost 3700 GWAS have suggested thousands of genetic risk variants and their corresponding biological function (110). Specifically, GWAS have been used to find the association between diseases such as type 2 diabetes, heart disease, neuropsychiatric disorders, various types of cancer plus common single nucleotide polymorphisms (SNPs) (111, 112). Next-generation sequencing (NGS) technologies allow the application of GWAS on a large scale, which, in conjunction with their drastically decreasing costs, have led to the aforementioned accumulation of data (113).

Personalized medicine

The emergence of big data in health sciences has paved the way for a more personalized approach in diagnostics, prognostics and therapeutics, as has been briefly mentioned above. Personalized medicine considers each patient's distinct characteristics and individualizes the medical approach (114). Specifically, it is thought that an individual’s molecular, physiological, behavioral and environmental exposure profile is responsible for the heterogeneity found in most diseases (115). The accumulation of data describing the aforementioned profile may help outline the completely distinct traits of an individual’s pathologic condition and how such a condition emerges, progresses, and is addressed. The final goal of personalized medicine is to provide the right treatment to the right person at the right time (116).

Thesis Overview

This thesis investigates nuclear receptors networks in relation to their role in the stress response system and the maintenance of homeostasis. The first section of this

thesis makes use of a literature study of transcription factors and their role in organism evolution and epigenetic modifications. This section is expected to display why a study on nuclear receptors, one of the largest superfamilies of transcription factors, may provide information useful for both clinicians and researchers. The next section of the thesis involves the analysis of a large dataset consisting of NR LBD structure and their corresponding ligands. Particularly, phylogenetic analysis is conducted to elucidate the evolutionary history of nuclear receptors further. The dataset was composed of more than 400 entries received from the Protein Data Bank (117). The results extracted provide novel insights on nuclear receptors' ligand-binding domain function. The third section of this thesis focuses on the glucocorticoid receptor, arguably the main regulator of the stress response system and makes use of data mining and semantics techniques to uncover novel information 'hidden' in the scientific literature and GWAS databases, such as GWAS Catalog (118), regarding the receptor's role in biological function and pathological conditions. The last part of the thesis attempts to analyze a more distinct dataset, particularly 3500 whole-genome sequences of Japanese individuals. Specifically, the glucocorticoid receptor is used as the basis to assemble a set of genes that participate in the stress response system, the regulation of various nuclear receptors action, the epigenetic regulation of gene transcription and drug metabolism. This set was then used to conduct a whole-genome study on this dataset in order to identify attributes that are specific to these individuals and may influence the way homeostasis is maintained in such populations.

The connection between evolution and transcription factors

Defining the mechanisms through which phenotypic variation is generated and its subsequent consequences are essential in providing novel information regarding the evolutionary history of life. Early on, geneticists considered genes as linearly arranged on chromosomes trait-causing elements (119). Later studies in developmental biology though, showcased several factors may influence gene function and thus gene action may be altered (120, 121). It is now accepted that a gene may express its encoded information in different ways, an ability that may have contributed in the emergence of phenotypic variation. Gene expression refers to the mechanisms through which the information encoded in a gene is used to generate a working gene product (122). Hence, gene expression allows organisms containing -mainly- the same DNA to showcase different cell types and functions (123). Gene expression is subject to multi-level regulation as a crucial mechanism in an organism's life (124). These regulating levels consist of the transcriptional level, the post-transcriptional level, the translational level, the post-translational level (125).

The above showcase that cell function and structure, although already known to be attributed to inherited genetic information, are also subject to influence by information not encoded in the DNA sequence. As stated briefly before, this information is called epigenetic information (123). Another definition for epigenetics includes both heritable alterations in gene expression and activity and stable, long-term alterations in the transcriptional potential of a cell that is not compulsorily heritable (126). As previously mentioned, there are multiple mechanisms associated with epigenetics that influence gene expression, such as DNA methylation, histone modifications (methylation, acetylation phosphorylation and ubiquitination) (127), chromatin remodeling, such as chromatin sliding (128), post-translation modifications (acetylation, amidation, glycosylation, methylation, phosphorylation) (129) and gene regulation led by various forms of regulatory RNA molecules, such as microRNAs (miRNAs) **(Figure 4)** (123, 130).

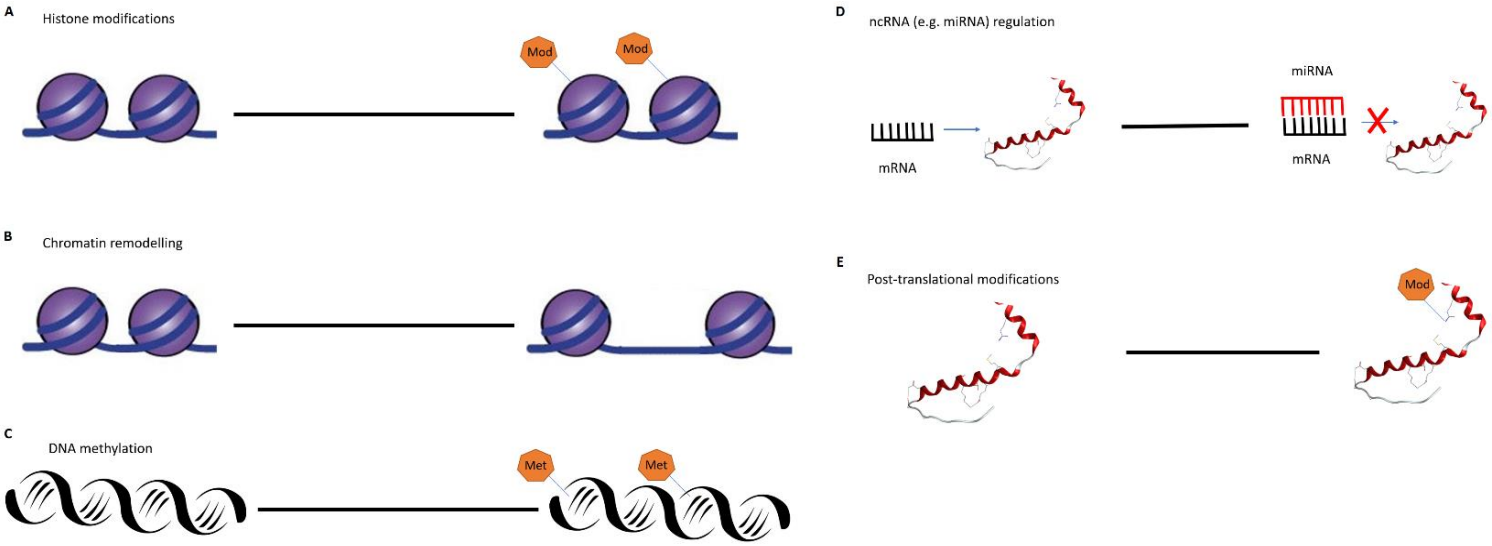

Figure 4. The various mechanisms of epigenetic modification. A) Histone modifications, such as methylation, acetylation phosphorylation, and ubiquitination B) Chromatin remodeling such as sliding

C) DNA methylation D) regulation by non-coding RNAs such as miRNAs E) Post-translational modifications such as acetylation, amidation, glycosylation, methylation, and phosphorylation.

Regulation of gene transcription is an essential component of gene activity in response to stimuli and tissue‐specific gene expression (131). Transcription factors (TFs) are the main regulators of gene transcription and are defined as proteins with the ability to bind specific DNA sequences to control gene expression (132). Different life forms make use of different methods for the initiation and regulation of transcription. Prokaryotes use a distinct RNA polymerase, while eukaryotes showcase multiple RNA polymerases (133). Bacteria make use of two distinct mechanisms for transcription initiation, a promoter-centric mechanism, in which a transcription factor interacts with a promoter and changes mentioned promoter's ability to bind the RNA polymerase, along with an RNA-centric mechanism in which a transcription factor interacts with the RNA polymerase to alter its promoter preference (134). In eukaryotes, several transcription factors interact with their respective DNA motifs, also known as response elements (REs) and recruit transcriptional cofactors (CoFs) to remodel the chromatin environment. These TFs can also promote the construction of a pre-initiation complex (PIC), which consists of general transcription factors (GTFs) and RNA polymerase II (RNAIIP) **(Figure 5)** (135). Finally, the transcriptional mechanism used by archaea may be described as a simplified version of the eukaryotic transcriptional mechanism. Archaeal TFs recruit the RNA polymerase to their respective DNA domain. Archaea make use of a transcriptional machinery that features additional RNA polymerase subunits and basal transcription factors that regulate transcription initiation and elongation (136).

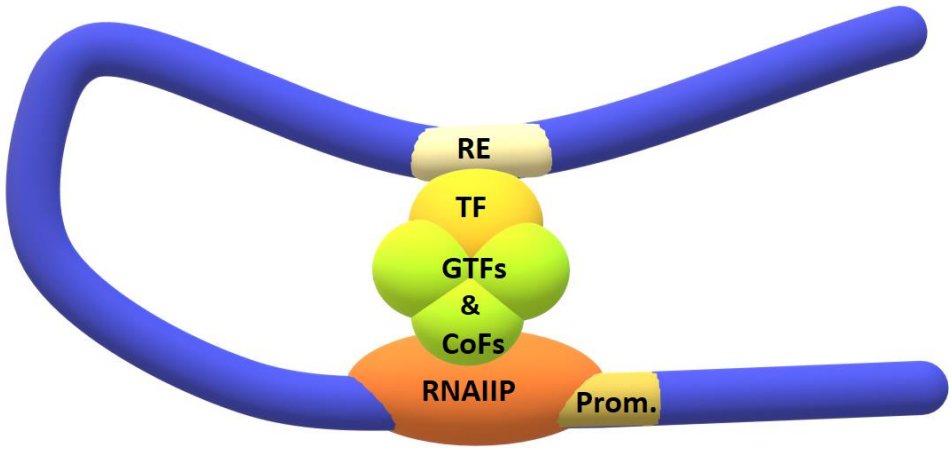

Figure 5. A schematic representation of a pre-initiation transcription complex in eukaryotes. RE: Response Element; TF: Transcription Factor; GTFs: General transcription factors; CoFs: Cofactors; RNAIIP: RNA-polymerase II; Prom: Promoter.

Transcription factors' activation is a complex process and may involve multiple intracellular signal transduction pathways, or direct regulation through binding of distinct molecules called ligands to a TF (137). Transcription factors regulate gene activity mostly through binding to distinct short DNA base pair patterns called motifs

or cis-regulatory elements (CREs) in downstream, intron, or upstream target gene regions. TFs also have the ability to interact with genomic locations that are distant from the primary DNA sequence (138). These locations are called gene regulatory regions (139). CREs contain promoters and DNA sequences called enhancers which participate in transcriptional activation and silencers which are DNA sequences that participate in transcriptional repression (140). The DNA-binding domain, which was mentioned as a structural region, is characteristic of all transcription factors. TFs display several structural motifs that recognize target DNA sequences like the homeodomain (HD), high mobility group box (HMG) and helix-turn-helix (HTH). These motifs can be used to classify transcription factors. TFs and DNA interaction is a quite complex process which is influenced by epigenetic modifications, cofactors' action, and the cooperative binding of other transcription factors (141). The above prove that gene regulation is a fundamental process in organism function. The emergence of various transcription factors throughout evolution and their respective effect on it may provide information regarding some of the most essential mechanisms driving life. This information can be further applied on nuclear receptors displaying their importance in biological function.

Transcription factors evolution among life domains

TF function is characterized by specific traits, the ability to identify and bind distinct short DNA sequences in regulatory regions and the ability to recruit or bind other proteins that also take part in transcriptional regulation (142). As a consequence, TFs' evolution is influenced by alterations in expression patterns, binding sites and binding partners (132). Additionally, since gene expression is somewhat inseparably linked with epigenetics, the evolution of epigenetics mechanisms is also associated with TFs' evolution (123). Lastly, events that are cornerstones of evolution are also part of transcription factors' evolution. Specifically, gene duplication and loss, the main drivers of the evolution of life, have an important role in TF evolution too. Specifically, duplication and deletion impact transcriptional regulatory networks by altering the number of TFs with distinct binding preferences (143-146). After the duplication of a TF gene, the resulting two copies are expected to be identical. The emerging TFs share the same sequences and consequently the same DNA-binding domain and thus bind the same target sequences. Possible consequent mutations in the DBD can lead a transcription factor copy to regulate a different target gene. Transcription factors display many differences on a lineage-specific level. The basal transcription machinery, although somewhat conserved in life, also exhibits diversification throughout evolution. The basal transcription machinery's subunit composition and size increase during evolution from 6 subunits in bacteria to 15 subunits in archaea. Moreover, eukaryotes, which are highly evolved, exhibit at least three different RNA polymerases (147). There are some apparent differences between eukaryotes and prokaryotes. Several DNA-binding domains are characteristic of some evolutionary lineages, like the ribbon-helix-helix domain, which is specific to bacteria and the Homeobox box, T-box, and C2H2-ZNfs domains, which are specific to eukaryotes (148). Furthermore, eukaryotic

transcription factors are somewhat longer than other eukaryotic proteins with other functions, while the opposite is true for prokaryotes. The reason for such a phenomenon may be that eukaryotic TFs contain a number of long intrinsic disordered segments that are necessary in order to participate in the formation of a multi‐protein transcription protein complex (149). Eukaryotes may also contain multiple repeats of the same DBD family in just one polypeptide chain. This ability may be a part of the mechanism eukaryotes use that increases the length and diversity of DNA binding recognition sequences while simultaneously using a limited number of DNA binding domain families (149).

Alterations in CREs have an impact on transcription factors' evolution and vice versa (150, 151). TFs can bind full promoter, enhancer, or silencer regions that feature multiple binding sites or a single DNA binding site. The factors which affect the emergence, disappearance and overall evolution of CREs include initial sequence distributions, which are biased towards the mutational neighborhood of strongly binding sequences, insertion and deletion mutational mechanisms, slippage processes, massive rearrangements of promoter regions and TF cooperation (152, 153). Insertion and deletion mutational mechanisms promote the slow emergence of binding sites out of a random sequence with the already sufficient genomic sequence from which sites can evolve and the potential co‐operativity between adjacent TFs helping accelerate such a process (152). It should also be stated that since a TF's interaction with its corresponding binding site is crucial in gene regulation, a mutation on either of them may interfere with their interaction and lead to deregulated gene expression. It is then no surprise that TFs' and CREs' evolution are interweaved, since such an association allows gene expression to remain functional. The co-evolutionary associations underlying TFs and CREs are based on the principle that a mutation in one of them may be compensated by a corresponding mutation in its interacting partner during evolution (154).

Transcription factors' co-operation

Individual prokaryote transcription factors have the ability to recognize long DNA motifs that themselves have the ability to define the genes they may regulate. On the other hand, organisms that have larger genomes are characterized by transcription factors that recognize short DNA sequences that cannot define unique genomic positions. Multicellular organisms also require molecular mechanisms that are complex and able to execute combinational processes for their development. These complex organisms have overcome the aforementioned hurdles by developing co‐operative recognition of DNA by multiple TFs during evolution. There are multiple mechanisms through which TFs can co-operate, with each mechanism dictating the specific details of the interaction. Some of these mechanisms include protein-protein interactions and indirect cooperation (155). The formation of functional dimers is a classic example of protein-protein interaction among TFs. Several eukaryotic TFs cannot bind DNA sequences as monomeric proteins and require physical interaction with an identical protein molecule or one within the same protein family to form functional dimers which have the ability to bind

targeted DNA sequences. It is thought that transcription factors used to function as monomers at first, a hypothesis supported by the fact that TFs can adequately bind target sequences as monomers in less complex organisms (156). Thus, it is suggested that some promoters which include symmetrical palindromic repeats of the DNA-recognition motif could have brought two or more copies of the same TF protein close. If coincidentally an interaction domain with only one interaction sequence emerged, then this would help establish the formation of the composite element on DNA since this complex would recognize a larger DNA motif (156). These evolutionary events could promote more relaxed evolutionary constraints on a transcription factor's DBD within a redundant duplicate gene and would allow the emergence of a still functional DNA-binding domain that binds with less affinity. If a duplicate gene diverts in such a way, it must from now on function as an obligate dimer. From then on, additional duplications and changes in specificity led to the emergence and diversification of the numerous TFs' dimerizing families (156). As has been already mentioned, TFs can also co-operate without direct protein-protein interaction. This TFs' function is achieved through a process known as indirect co-operativity or collaborative competition, where a cohort of TFs collectively competes with the same histone octamer in order to access the underlying DNA (157). Indirect co-operativity emerges from the close juxtaposition of binding sites for arbitrarily chosen transcription factors (158). Thus, collaborative competition may have an essential role in the evolution of gene regulatory modules since molecules that undergo combinatorial regulation can be assembled from randomly selected components and show no requirement for coevolution. Possible coevolution of the required partners may enhance co-operativity through the aforementioned protein-protein contacts or through bridging proteins and therefore increase the magnitude of combinational control (158).

Coactivators, corepressors, and post-translational modifications

Several coactivator and corepressor proteins that regulate TFs partake in multi-subunit coregulator complexes and present enzymatic activities (159). Such complexes can regulate TF activity as a response to stimuli through post-translational modifications (PTMs) (160). Modification enzymes interact directly with transcription factors and modify distinct residues of the TF protein and change subcellular localization, interaction with additional cofactors, stability, along with other transcriptional activities (161). Some of these post-translational modifications these enzymes carry out are acetylation, glycosylation, methylation and phosphorylation (160). It is quite possible that post-translational modifications of transcription factors, histones, or RNA polymerase II and its associated proteins at the pre-initiation complex participate in enhancer-core-promoter communication and possibly, in the combinatorial regulation of transcription activation (162). Therefore, it is not surprising that, in the past decade, research has identified connections between novel post-translational modification sites within TFs and the emergence and evolution of new features (142). A distinctive example is pregnancy evolution in mammals where amino-acid differentiation in the TF CCAAT/enhancer-binding protein beta (CEBPB) alter the way in which this TF

responds to cyclic AMP/protein kinase A (cAMP/PKA) signaling. Such differences in the amino-acid sequence alter the location of key phosphorylation sites and potentially change the response of CEBPB to phosphorylation from activation to repression (163).

Expression patterns in transcription factors

TFs in eukaryotes display functional differences among different species and paralogs, a characteristic that demonstrates their ability to evolve new functions (164). Eukaryotes showcase five groups of transcription factors with specific expression patterns that have emerged through periodic expansion in the transcription factors' repertoire. These groups are split into those that are only present in primates, those that are mostly found in mammals, or metazoan and those found in the majority of eukaryotes, including yeast (165). A mechanism that can promote such changes in TF function is tissue‐specific gene expression. Tissue specificity allows the minimization of mutations' pleiotropic effects that possibly lead to the gain of novel regulatory links via transcription factor evolution. Additionally, tissue specificity limits the loss of function mutations' effects, which break regulatory links (164). Combining the study of the expansion that occurred in the TF repertoire with the study of tissue‐specific mechanisms' evolution can help elucidate the association between TFs and tissue specificity. The previously mentioned expansion appears to have occurred unevenly for TFs containing different types of DBDs. Some DNA-binding domains showcase a rapid expansion through evolution, while others exhibit no significant expansion since their emergence. This expansion in TF ơrepertoire may have provided evolution with the tools needed to create or modify different expression patterns for transcriptional factors -including tissue‐specific ones- by duplication and subsequent promoter divergence (165).

Histone modifications and transcription factors

DNA wraps around histone octamers in human cells. The resulting complex is termed 'nucleosomal core particle' (166). Histone proteins exhibit tails, which feature residues that can be post‐translationally modified and later influence transcription. Changes in histone modification patterns that characterize TF binding sites regulate the aforementioned influence on transcription. Moreover, TFs with evolutionary-related DNA binding domains are thought to sample putative binding sites whose environment displays similar histone modification patterns (167).

MicroRNAs and transcription factors

MicroRNAs, also abbreviated as miRNAs or miRs, are small non-coding RNA molecules that can regulate gene expression (168). Both transcription factors and miRNAs have an essential role in gene regulatory network evolution (169). Gene expression appears to be regulated by transcription factors at the transcriptional

level and miRNAs at the post-transcriptional level, with both TFs and miRNAs being able to regulate each other. The interaction between TFs and miRNAs provides distinct constraints and functional alterations for the evolution of gene expression regulatory networks. It is not surprising then, that these two regulators display some form of coevolution (170). Specifically, coevolution appears in TF and miRNAs pairs that are associated through transcriptional activation signals but not in pairs that are associated through transcriptional repression signals. The above may be due to the fact that TFs that trigger miRNA expression may afterward act in tandem with the activated miRNAs, while TFs that repress the expression of specific miRNAs will not act with them and therefore be under their evolutionary influence (170).

Focusing on nuclear receptors' evolution

Transcription factors' activation method has a tremendous effect on their evolution. Ligands in ligand-activated transcription factors, such as nuclear receptors, are expected to have an integral part in the evolution of these regulatory proteins. NR ligands are product intermediates of a wide variety of metabolic pathways.
Therefore, these ligands have been evolutionary set through genetic modulations on the components of particular metabolic pathways but not through single gene modifications. This evolutionary characteristic of ligands implies that nuclear receptors' evolution is heavily influenced by the ever‐growing interaction among gene networks and not exclusively by a group of distinct genes (171). This specific characteristic of NR evolution makes its research a complex procedure. On the other hand, nuclear receptors are the only TFs that present a direct link between the metabolic environment and gene regulation, while may have additionally played an essential role in the diversification of animals as multicellular heterotrophs (172). Thus elucidating NR evolution, particularly the way ligands interact with their corresponding receptor, is the desired result since such information has implications not only on the proteins themselves but on their ligands and intricate physiological and metabolic mechanisms, including gene regulation.

An in-depth view on nuclear receptors' LBD structure and evolution

Studies on the origin of nuclear receptors suggest that they did not show a high affinity for a specific ligand initially, with mentioned ability being acquired later through evolution. It is thought that the first nuclear receptor was able to bind different ligands with low affinity and selectivity (172). The existence of a single ancestral NR that can bind different ligands with distinct biological activities is in accordance with the selective NR modulators concept. Selective NR modulators refer to molecules that can selectively activate a nuclear receptor in a tissue or target-specific fashion. The aforementioned concept provides a theoretical foundation for the ability of several currently existing nuclear receptors to bind different ligands with different selectivity (171). Therefore, studying the structure and sequence of the LBD domain throughout evolution in conjunction with their corresponding ligands can provide novel information regarding NR evolution and function.

Methods of study

In order to collect information regarding NR LBD, a search was performed on the RSCB Protein Data Bank (PDB) database for amino-acid sequences that have been associated with the ligand-binding domain of nuclear receptors (117). Any sequence that responded to the query but did not feature a ligand-binding domain was removed from the dataset through the use of regular expression techniques and local alignments with reference sequences. Roughly 400 NR LBDs were protein sequences, and structures were collected from a variety of species **(Table 2)**.

Table 2. Structures used during the current study

Nuclear Receptor	**PDB Structure**
Ancestral Corticoid Receptor	2Q1H, 2Q3Y, 4FNE, 4LTW
Glucocorticoid Receptor (NR3C1)	4CSJ, 1NHZ, 5NFT, 1M2Z, 3CLD, 3E7C, 4LSJ, 3BQD, 5UC3, 4MDD, 4P6X, 3MNE, 3H52, 3MNO, 4P6W, 3MNP, 5UC1 (GRβ)
Ancestral Glucocorticoid Receptor	3GN8, 5UFS, 4E2J, 3RY9
Progesterone Receptor (NR3C3)	3G8O, 1SQN, 1E3K, 1A28, 1SR7, 3KBA, 4OAR, 2W8Y
Androgen Receptor (NR3C4)	3RLL, 5JJM, 2OZ7, 5T8E, 3L3X, 2Q7K, 5VO4, 3RLJ, 4OGH, 2Z4J, 1XOW, 1T5Z, 2AX6, 2HVC, 1I38, 4K7A, 4QL8, 2AM9, 4OEA, 2AX9, 1I37, 1E3G, 1XJ7, 1GS4, 5CJ6, 1T73, 5V8Q, 4OJB, 2AX8, 1Z95
Estrogen Receptor alpha (NR3A1)	1PCG, 5T92, 5ACC, 5AAV, 5W9D, 1QKT, 5FQP, 1UOM, 5AAU, 2YAT, 6CHZ, 5AK2, 3UU7, 3Q97, 4MG8, 1ZKY, 4MG5, 3UUA, 5DXG, 6CBZ, 5T1Z, 4MG7, 2QZO, 2QA8,3D24, 3HM1, 1ERR, 3Q95, 3UUC, 4IU7, 2QA8, 2OCF, 2P15, 5DX3, 4Q50, 4ZN9, 5HYR, 2BJ4, 5GS4, 2IOG, 5DXB, 1SJ0, 5DI7, 3HLV, 2JFA, 4PP6, 4DMA, 2QXS, 4Q13, 1GWQ, 1ERE, 2IOK, 1G50, 2PJL, 3L03, 1L2I, 4PXM, 1GWR, 3ERD,

	1QKU, 3DT3, 1X7E, 1A52, 1XB7, 3OS8, 5TN9, 4XI3, 4N1Y
Estrogen Receptor beta (NR3A2)	2FSZ, 1YY4, 1ZAF, 2NV7, 3OLL, 4ZI1, 1X76, 2YLY, 1U9E, 1QKM, 2GIU, 1NDE, 1L2J, 2I0G, 1QKN, 2J7X,
Ecdysone Receptor (EcR protein and Ultraspiracle Protein) (NR1H1 and NR2B4)	1G2N, 1R1K, 1R20, 2NXX, 4OZT, 1Z5X, 4OZR, 1HG4, 2NXX
TLX receptor (NR2E1)	4XAJ,4XAI
Steroidogenic factor 1 (NR5A1)	3F7D, 1YMT, 1YP0, 4QJR, 1ZDT, 1YOW,
Nur77 (NR4A1)	4RZF, 1YJE, 2QW4, 4JGV, 4REF,4KZJ, 4RZG, 3V3E, 4KZM,4RZE
Nurr1 (NR4A2)	1OVL,
Liver receptor homolog-1 (NR5A2)	4ONI, 4RWV, 1YUC, 1P5K, 1ZDU, 5L11, 5SYZ, 4PLD, 4DOS, 3PLZ, 1YOK, 3TX7, 4IS8
Photoreceptor cell-specific nuclear receptor (NR2E3)	4LOG
Hepatocyte nuclear factor 4 alpha (NR2A1)	1M7W, 4B7W, 3FS1, 1PZL, 1LV2, 4IQR
Peroxisome proliferator-activated receptor alpha (NR1C1)	1I7G, 4BCR, 2P54, 2NPA, 3G8I, 5AZT, 2REW, 1KKQ, 3SP6, 2ZNN, 5HYK, 1K7L, 3ET1
Peroxisome proliferator-activated receptor gamma (NR1C3)	2Q59, 4PRG, 2PRG, 2ZVT, 3U9Q, 3CS8, 5GTN, 4O8F, 3OSI, 6ENQ, 2I4J, 3CWD, 4L96, 3R8A, 3ADS, 5JI0, 1KNU, 1FM6, 3PBA, 3WJ4, 2Q59, 3B0Q, 3VN2, 5Y2O, 1K74, 3LMP, 2VV1, 1RDT, 1ZEO, 4EM9, 2VSR, 2VV4, 4R06, 5LSG, 3T03, 3S9S, 1WM0, 3ET0, 2OM9, 4R2U, 4R6S, 1I7I, 3PRG, 4OJ4, 1NYX, 3ET3, 2HFP, 3BC5, 3DZU
Peroxisome proliferator-activated receptor beta/delta (NR1C2)	2GWX, 1GWX, 3SP9, 2B50, 5U3Q, 2AWH, 2Q5G, 1Y0S, 5U3T, 3TKM, 2J14, 2XYG, 2ZNP, 3ET2, 3GZ9
Retinoic acid receptor alpha (NR1B1)	3A9E, 5K13, 1DKF, 4DQM, 3KMR
Retinoic acid receptor beta (NR1B2)	1XDK, 1XAP, 4JYG, 5UAN
Fushi tarazu factor 1 (NR5A3)	2XHS
Rev-Erb beta (NR1D2)	2V0V, 4N73, 2V7C, 3CQV
Liver X-Receptor alpha (NR1H3)	2ACL, 3FC6, 1UHL, 3IPQ
Liver X receptor beta (NR1H2)	3L0E, 4DK8, 1P8D, 1PQ6, 1UPV, 5HJP, 4RAK, 5I4V, 5KYA, 4NQA, 4DK7
Vitamin D receptor (NR1I1)	3A2I, 3M7R, 3AZ1, 3B0T, 1S0Z, 1DB1, 5V39, 5GT4, 3P8X, 4RUJ, 4FHH, 2HBH, 3A2J
Farnesoid X receptor (NR1H4)	3GD2, 3P88, 4OIV, 3HC6, 3P89, 3BEJ, 3DCT, 3HC5, 3L1B, 3FLI, 5WZX, 1OSH, 4QE6, 5IAW, 3RUT, 3FXV, 1OSV, 1OT7, 5Q0L, 4WVD
Thyroid hormone receptor alpha (NR1A1)	3UVV, 1NAV, 3JZB, 3HZF, 2H77, 2H79
Thyroid hormone receptor beta (NR1A2)	2J4A, 1Q4X, 4ZO1, 1R6G, 1NUO, 1NQ0, 2PIN, 1NQ1, 1BSX, 1N46, 1XZX, 1NQ2, 3IMY, 3JZC, 1NAX, 3D57, 3GWS

Testicular receptor 4 (NR2C2)	3P0U
Retinoid X receptor in *Biomphalaria Glabrata*	1XIU
Retinoid X receptor in *Polyandrocarpa Misakiensis*	2Q60
Retinoid X receptor gamma (NR2B3)	2GL8
RAR-related orphan receptor alpha (NR1F1)	4S15, 1N83
RAR-related orphan receptor beta (NR1F2)	1NQ7, 1K4W, 1N4H,
RAR-related orphan receptor gamma (NR1F3)	5IXK, 3L0L, 5IZ0, 5EJV, 5YP6, 4WPF, 4NIE, 5AYG, 4S14, 4WLB, 5X8Q, 5UFO, 5APK, 5NTI, 6B30, 5C4O, 5ETH, 5LWP, 5W4V, 6BR3, 3KYT, 5W4R, 6CVH, 5NTP, 5K38, 4NB6, 6BN6, 5M96, 5APH, 4ZRJ, 4YMQ

The MATLAB Bioinformatics Toolbox was used to perform a multiple sequence alignment, specifically through making use of the progressive multiple alignment method and a guide tree (173, 174). Pairwise distances were estimated after pairwise alignment with the Gonnet scoring matrix and counting the proportion of sites at which sequence pairs are different (175). The guide tree was computed through the neighbor-joining method and assuming equal variance and independence of evolutionary distance estimates. Visualization of consensus sequences was achieved through the use of the Jalview platform and based on several multiple sequence alignment results and parameters such as amino acids conservation and quality (176). A more in-depth alignment focusing on characteristic steroid hormone receptors was conducted using the Molecular Operating Environment (MOE) (177).

A comprehensive structural and functional analysis of NR LBDs was achieved by comparing different LBD structures by superposing the structures and calculating the root mean square deviation (RMSD). The structural superposition method, as made possible by the MATLAB Bioinformatics Toolbox, was used to compare the structures of the NR LBDs (178). Structural superposition computes and applies a linear transformation to superpose the coordinates first structure's atoms to the coordinates of the second structure's atoms. Single chains' alpha carbon atom coordinates for each structure are taken into account for computing the linear transformation. The structural similarity matrix is shown using MATLAB in five different colors (blue for a range of 0 to 1, light blue for a range of 1.1 to 2, light gray for a range of 2.1 to 3, orange for a range of 3.1 to 4.9 and red for values ≥5).

MOE was used to gain a more thorough view of steroid hormone receptor structures (179-181). All steroid hormone receptors found in the original dataset and their corresponding ligands were studied. The receptor-ligand interactions were the main focus of this study section, where each PDB entry was examined for ligand interaction using the MOE ligand interaction function. This function helped display the LBD amino acids that interact with their corresponding ligands or co-activators.

Lastly, information from the NCBI conserved domain database was extracted and assigned to the MOE results.

The unweighted pair group method (UPGMA), as made possible by the MATLAB Bioinformatics Toolbox, along with a specific hybrid matrix of pairwise distances, was used to perform a specialized phylogenetic analysis (182-185). This specific matrix combines information from the distance matrix of the multiple sequence alignment and the RMSD matrix of the structural analysis. Element by element matrices proliferation is used to calculate mentioned combined matrix (186, 187). This method helps cluster proteins that are less similar on a sequence level but are more conserved on a structural level. Lastly, the resulting phylogenetic tree was visualized using the MEGA software, specifically the radiation option, with the final clusters being separated by different colors (188).

It is thought that studying chemical similarity may help predict chemical compounds' properties, cluster chemicals and, more importantly, be the basis for a thorough functional analysis. Similarity calculation between any two molecules can be accomplished by comparing their respective molecular fingerprints (189). Such fingerprints consist of a molecule's structural information that has been encoded as a series of bits. The Tanimoto coefficient is the most popular method of comparing chemical structures' similarity, with mentioned structures being represented by molecular fingerprints (190).

All ligands that have been co-crystallized with SHRs in the dataset being studied were extracted **(Table 3)**. Specifically, ninety-four ligands unique to steroid hormone receptors were collected and were compared in order to identify possible similarities. These ligands' structures were compared using the Tanimoto coefficient algorithm (191). The Tanimoto coefficient varies from 0, when the fingerprints have zero bits in common, to 1, when the fingerprints are identical. The results of this comparison were saved in a chemical-specific similarity matrix. The chemical-specific similarity matrix was visualized using MATLAB in 4 different colors (black for a coefficient that ranges from 0 to 0,59; purple for a coefficient that ranges from 0,6 to 0,69; light blue for a coefficient that ranges from 0,7 to 0,89; and blue for a coefficient that ranges from 0,9 to 1)

Table 3. A list featuring all steroid hormone receptors ligands that have been studied and their corresponding receptor

Ligand /Molecular Formula/ CID	Order number	Receptor	Positions of Int	PDB IDs
8W8 C_{25} H_{21} F_4 N_3 O_3 24825740	1	GR	Asn564, Gln570, Gln642	5NFT
486 C_{29} H_{35} N O_2 55245	2	GR AncCor	Leu563, Leu566, Gln570, Arg611, Gln642, Cys736 Gln39, Arg80, Cys205,	1NHZ,5UC3,3H52,5UC1, (4LTW)

			Tyr208	
DEX C_{22} H_{29} F O_5 5743	3	GR AncCor	Leu563, Asn564, Gln570, Arg611, Phe623, Gln642, Met646, Cys736, Thr739 Leu29, Asn33, Gln39, Arg80, Leu111, Cys205, Thr208	1M2Z, 3NE, 3MNO, 3MNP, 3GN8
NN7 C_{25} H_{27} F N_4 O_2 S 16666386	4	GR	Asn564, Gln570, Met604, Leu608, Phe623, Cys736	4CSJ
GW6 C_{27} H_{29} F_3 O_6 S 9854489	5	GR	Asn564, Arg611, Cys736	3CLD
866 C_{23} H_{21} Cl_2 F_4 N_5 O_3 25058139	6	GR	Asn564, Gln570, Met604, Gln642, Cys736	3E7C
LSJ C_{25} H_{25} N O_4 S 72710581	7	GR	Met560, Asn564, Cys736, Thr739	4LSJ
29M C_{28} H_{32} F N_3 O_3 S 86280440	8	GR	Met560, Leu563, Asn564, Cys736	4MDD
HCY C_{21} H_{30} O_5 5754	9	GR	Asn564, Gln570, Arg611, Thr739	4P6X
DAY C_{30} H_{36} N_2 O_4 3032474	10	GR	Met560, Asn564, Phe623, Gln642	3BQD
MOF C_{27} H_{30} Cl_2 O_6 441336	11	GR AncCor PR	Asn564, Gln570, Arg611, Cys736 Asn33, Gln39, Arg80 Asn719, Cys891	4P6W, 4E2J, 1SR7
1TA C_{24} H_{31} F O_6 6436	12	AncCor	Leu29, Asn33, Gln39, Thr208	5UFS
1CA C_{21} H_{30} O_3 6166	13	AncCor MR	Asn33, Gln39, Arg80, Cys205, Thr208 Asn770, Gln776, Cys942, Thr945	3RY9, 2Q3Y, 4FNE, 2ABI
AS4 C_{21} H_{28} O_5 5839	14	AncCor MR	Asn33, Gln39, Arg80, Cys205, Thr208 Asn770, Gln776, Cys942, Thr945	2Q1H, 2AA2
SNL C_{24} H_{32} O_4 S	15	MR	Asn770, Gln776,	3VHU

5833			Arg817	
WFF C_{18} H_{11} F_2 N O_4 72163477	16	MR	Asn770, Cys849, Thr945	3WFF
LD1 C_{18} H_{13} N_5 O_2 S 54751696 LD2 C_{18} H_{15} N_5 O_2 S 54751697	17	MR	Leu769, Asn770, Met807, Ser811, Met845, Thr945 | Ala844, Leu848	3VHV
30X C_{15} H_{15} F_6 N_3 O 11326074	18	PR	Leu715, Met756, Arg766, Cys891	3G8O
NDR C_{20} H_{26} O_2 6230	19	PR	Cys891	1SQN
R18 C_{19} H_{24} O_2 261000	20	PR AR	Asn719, Gln725, Arg766, Cys891 Asn705, Arg752, Thr877	1E3K, 1XOW, 1E3G
STR C_{21} H_{30} O_2 5994	21	PR	Asn719, Gln725, Arg766, Cys891	(4LTW), 1A28
WOW C_{20} H_{22} Cl N_3 O_2 S 16661548	22	PR	Arg766, Cys891, Thr894	3KBA
2S0 C_{30} H_{37} N O_4 130904	23	PR	Gln725, Arg766, Cys891, Thr894	4OAR
RLL C_{23} H_{16} F_3 N_3 O_3 51346204	24	AR	Leu704, Asn705, Trp741, Arg752	3RLL
DHT C_{19} H_{30} O_2 10635	25	AR	Asn705, Gln711, Glu793, Thr877	5JJM(a+b), 3L3X, 2Z4J, 1T5Z, 1I38, 4K7A, 4OEA, 1XJ7, 1T73
CA4 C_{24} H_{29} Cl O_4 9880	26	AR	Asn705, Gln711	2OZ7
77U C_{13} H_{15} Cl N_2 O 59370500	27	AR	Asn705, Gln711, Met745, Arg752, Thr877	5T8E
TES C_{19} H_{28} O_2 6013	28	AR	Asn705, Thr877	2Q7K, 2AM9
9FG C_{12} H_9 F N_2 O 132471744	29	AR	Gln711, Met745	5VO4
RLJ C_{19} H_{14} F_3 N_3 O_3 11326715	30	AR	Leu704, Asn705, Gln711, Met745, Thr877, Met895	3RLJ
HFT C_{11} H_{11} F_3 N_2 O_4 91649	31	AR	Leu704, Asn705, Met895	4OGH, 2AX6
LGD C_{14} H_9 F_9 N_2 O 11560224	32	AR	Asn705, Gln711, Arg752	2HVC
MXD C_9 H_{15} N_5 O 4201	33	AR	Asn705, Thr877	4K7A
JAD C_{15} H_{16} Cl N_3 O_2 71239417	34	AR	Asn705, Met745, Arg752, Thr877	4QL8

BHM C_{11} H_{10} Br F_3 N_2 O_4 5287785	35	AR	Leu704, Asn705, Thr877	2AX9
ZK5 C_{21} H_{29} F O_5 31378	36	AR	Asn705, Gln711, Arg752 (+++mutations His701, Ala877)	1GS4
51Y C_{14} H_{17} Cl N_2 O 71543393	37	AR	Asn704, Gln711, Met745, Arg752	5CJ6
97A C_{14} H_{13} F_3 N_2 O_2 59556974	38	AR	Asn704, Thr877	5V8Q
198 C_{18} H_{14} F_4 N_2 O_4 S 56069	39	AR	Leu704, Asn705, Gly708, Gln711, Met742, Arg752, Thr877	4OJB, 1Z95
FHM C_{17} H_{14} F_4 N_2 O_5 5288215	40	AR	Leu704, Asn705	2AX8
EST C_{18} H_{24} O_2 5757	41	ERa ERb	Met343, Glu353, His524 His475	1PCG, 1QKT, 5DXG, 6CBZ (a+b), 2OCF, 5DX3(a+b), 5HYR, 5GS4, 5DXB, 1ERE, 1G50, 4PXM, 1GWR, 1QKU, 1A52, 3OLL
77W C_{25} H_{22} F N O_3 118166742	42	ERa	-	5T92
KE9 C_{25} H_{25} F_3 N_2 O_2 86287635	43	ERa	Leu346, Met421, Val533	5ACC
GW5 C_{25} H_{22} O_2 5288494	44	ERa	-	5AAV(a+b)
9XY C_{25} H_{27} N O_2 10090750	45	ERa	Thr347, Asp351, Val533	5W9D
GQD C_{23} H_{27} N O_3 127034153	46	ERa	Met421, Val533, Val534	5FQP
PTI C_{28} H_{32} N_2 O_2 448915	47	ERa	Asp351	1UOM
F3D C_{31} H_{36} N_4 O_2 134519316	48	ERa	Asp351, Glu353, Leu387, Arg394, Cys530	6CHW, 6CHZ
XBR C_{20} H_{19} Cl N_2 O_2 91936962	49	ERa	Gly521, Val533	5AAU
EEU C_{35} H_{41} N_3 O_{10} 11614456	50	ERa	Glu423	2YAT
85Z C_{26} H_{19} F O_5 91668558	51	ERa	Phe404, Val534	5AK2
2OH C_{15} H_{16} O_2 6623	52	ERa	-	3UU7(a+b)

Q97 $C_{24}H_{24}O_3$ 46205471	53	ERa	-	5T1Z
27J $C_{18}H_{26}O_5$ 2999413	54	ERa	Glu353, His524	4MG8
689 $C_{18}H_{24}O_3$ 6857699	55	ERa	Met421, His524	1ZKY
27E $C_{10}Cl_{10}O$ 299	56	ERa	Met343, Leu346, Met421	4MG5
0CZ $C_{15}H_{10}F_6O_2$ 73864	57	ERa	Glu353	3UUA(a+b)
GEN $C_{15}H_{10}O_5$ 5280961	58	ERa ERb	His524 Met336, His475	2QA8(a+b), 1QKM
J3Z $C_{18}H_{22}O_2$ 5870	59	ERa	Glu353, Gly521	3HM1
ESL $C_{18}H_{24}O_3$ 5756	60	ERa	Glu353, Met421, His524	3Q95(a+b)
RAL $C_{28}H_{27}NO_4S$ 5035	61	ERa	Asp351, Glu353, Lys362, Val533, Glu542	1ERR, 2QXS, 2JFA_b
ZTW $C_{14}H_{10}O_2S$ 445920	62	ERa	-	1GWQ
0D1 $C_{14}H_{10}Cl_2O_2$ 84677	63	ERa	-	3UUC
1GM $C_{16}H_{13}F_3N_2O_2$ 135566804	64	ERa	Leu346	4IU7
OHT $C_{26}H_{29}NO_2$ 449459	65	ERa ERb	Asp351, Val533 Asp303, Gln327	2BJ4, 4Q50, 2FSZ
OBH $C_{24}H_{22}O_6S$ 51006494	66	ERa	Glu353, Gly521	4ZN9
IOG $C_{33}H_{39}N_3O_3$ 16750039	67	ERa	Asp351, Glu353, Gly521, Cys530, Lys531	2IOG
E4D $C_{27}H_{29}NO_4S$ 448577	68	ERa	Asp351, Glu353, Leu387, Cys530	1SJ0
5CQ $C_{17}H_{24}O_2$ 50940842	69	ERa	Met343, His524	5DI7
J2Z $C_{18}H_{22}O_3$ 115116	70	ERa	Met421, Gly521	3HLV
STL $C_{14}H_{12}O_3$ 445154	71	ERa	Glu353, Arg394	4PP6
0L8 $C_{17}H_{13}BrO_3$ 11588238	72	ERa	Met343, His524, Leu525	4DMA
IOK $C_{26}H_{26}N_2O_2$ 16750040	73	ERa	Met343	2IOK

047 C_{23} H_{28} N_2 16122612	74	ERa	Glu331	2PJL
4OH C_{18} H_{24} O_4 27125	75	ERa	His524	3L03
ETC C_{22} H_{24} O_2 446849	76	ERa ERb	Asp321, Trp360, Lys362, Val364 Leu339	1L2I, 1L2J
DES C_{18} H_{20} O_2 448537	77	ERa	-	3ERD
369 C_{23} H_{18} O_4 24892830	78	ERa	Glu353, Phe404, His524	3DT3
244 C_{16} H_{11} N O_3 656953	79	ERa	Glu353, His524	1X7E
	80	ERa		
KN0 C_{21} H_{15} F_3 N_2 O_2 135430624	80	ERa	Leu346, Glu353, Leu387	3OS8(a+c)
7EC C_{31} H_{32} Br N O_6 S	81	ERa	Asp351, Glu353, Gly521, Asn532, Val533	5TN9
29S C_{30} H_{34} N_2 O_3 154257	82	ERa	Asp351, Glu353, His524, Asn532 Val533	4XI3
KN1 C_{17} H_{13} F_3 N_2 O_2 135461982	83	ERa	-	2QZO
EZT C_{27} H_{29} F_3 O_2 15485192	84	ERa	His524	2P15
27H C_{22} H_{30} O_4 354654	85	ERa	Glu353	4MG7(a+b)
4NA C_{16} H_{11} Cl O_2 6102690	85	ERb	Glu305	1YY4
789 C_{15} H_9 Br O_3 10286462	86	ERb	His475	1ZAF
555 C_{17} H_{13} N O_2 135440536	87	ERb	-	2NV7
KB0 C_{22} H_{18} O_2 58027337	88	ERb	Glu305, His475	4ZI1
697 C_{15} H_9 N O_3 656952	89	ERb	Glu305, His475	1X76
SU4 C_{18} H_{21} N O_4 S 53483961	90	ERb	-	2YLY
397 C_{14} H_{10} O_3 656936	91	ERb	Glu305	1U9E
FBR C_{17} H_{19} Br O_2 11987846	92	ERb	Glu305	2GIU
MON C_{24} H_{29} Cl N_6 O S	93	ERb	Met295, Asp303	1NDE

9957008			Glu305, Leu476, Met479	
I0G C_{18} H_{18} O_3 10286159	94	ERb	-	2I0G

Additionally, a literature review was conducted to obtain all known mutations found in NR LBDs as of January 2019 **(Table 4)**. A study of the naturally occurring mutations on NR LBDs can highlight regions of possible evolutionary importance, while studying mutations residing on the motifs found by the aforementioned structural analysis may provide more concrete information about mentioned motifs' role in receptor function.

Table 4. Naturally occurring mutations found on nuclear receptors' ligand-binding domains and their position on the multiple alignment conducted

Serial number	Receptor	Mutation	Multiple sequence alignment position	PubMed ID of corresponding mutations
1	GR	L773P, 2bp del CT, 2bp del TG	647	8316249 23076843 19933394
2	GR	L753F	612	8316249
3	GR	I747M	602	12050230
4	GR	F737L	591	17635946
5	GR	V729I	583	7683692
6	GR	H726R	580	26031419
7	GR	R714Q	567	20335448
8	GR	G679S	522	11589680
9	GR	L672P	515	27120390
10	GR	D641V	484	1704018
11	GR	612 1bp del	429	20861124
12	GR	588 ins TTAC	405	27211791
13	GR	V575G	392	24483153
14	GR	V571A	388	11932321
15	GR	I559N	376	8863343
16	GR	T556I	373	21362280
17	GR	Q501H	318	29444898
18	AR	F917L	648	22334387
19	AR	P914S	645	22334387
20	AR	L908F	624	22334387
21	AR	P905H/S	621	22334387
22	AR	P893S	609	22334387
23	AR	V890M	602	22334387
24	AR	S889 del (no immunreactiv.)	601	22334387
25	AR	T878A	589	11906285

26	AR	H875Y	587	22334387
27	AR	I870M	582	22334387
28	AR	V867M/L	579	22334387
29	AR	G821A	522	22334387
30	AR	P818A	519	22334387
31	AR	L813P	514	22334387
32	AR	M808V/T/R	509	22334387
33	AR	E804K	505	22334387
34	AR	F795S	496	22334387
35	AR	M788V	489	22334387
36	AR	R787X	488	22334387
37	AR	V786 2bp del	487	22334387
38	AR	C785Y	486	22334387
39	AR	Y782D	483	22334387
40	AR	M781I	482	22334387
41	AR	R775H/C	475	22334387
42	AR	E773G	473	22334387
43	AR	L769M	444	22334387
44	AR	P767S	442	22334387
45	AR	A766T	441	22334387
46	AR	F765L	440	22334387
47	AR	Y764C	439	22334387
48	AR	F755L/V	430	22334387
49	AR	R753Q/X	428	22334387
50	AR	W752R/X	427	22334387
51	AR	M750V	425	22334387
52	AR	M746T	421	22334387
53	AR	G744E/V	419	22334387
54	AR	M743I/V	418	22334387
55	AR	W742C	417	22334387
56	AR	Y740D	415	22334387
57	AR	D733Y	408	22334387
58	AR	G725D/S	400	22334387
59	AR	P724S	399	22334387
60	AR	W719X	394	22334387
61	AR	L713F	388	22334387
62	AR	R711T	386	22334387
63	AR	N706S	381	22334387
64	AR	S704C/G	378	22334387
65	AR	L702H	377	22334387
66	AR	L701M	376	22334387
67	AR	N693 del	368	22334387
68	AR	G689X	364	22334387
69	AR	V685I	360	22334387
70	AR	I681N	356	22334387
71	AR	L678P	353	22334387

72	ERa	D538G	607	25838462
73	ERa	Y537N/C/S	602	26122181
74	ERa	L536P/R/Q	601	25838462 26122181 26183887
75	ERa	P535H	600	26122181
76	ERa	V534E	599	26122181
77	ERa	K531E	596	15583021
78	ERa	R503W	567	24398047
79	ERa	S463P	524	26122181
80	ERa	432del- 437stop	493	15583021
81	ERa	M427I/L429M	488/490	15475371
82	ERa	G415V	471	15583021
83	ERa	411fsh- 418stop	472	16713253
84	ERa	G400V	434	15583021
85	ERa	M396V	430	16713253
86	ERa	E380Q	414	26122181
87	ERa	E353V	387	16713253
88	ERa	344insCyst	378	26183887
89	ERa	S309F	343	16713253
90	ERa	K303R	337	26183887
91	PPARa	V227A	330	16288935
92	PPARg	P467L	607	28208577
93	PPARg	H449L	586	26756202
94	PPARg	R397C	530	28208577
95	PPARg	D396N	518	17766367
96	PPARg	F360L	482	25004973
97	PPARg	R357A/X	479	28208577 18713822
98	PPARg	Y327X	423	6412238
99	PPARg	K319X	414	10394368
100	PPARg	312fs-315stop	407	18713822
101	PPARg	L311X	406	23393388
102	PPARg	V290M	385	28208577
103	PPARg	Q286P	381	28208577
104	PPARg	R280P	375	29622583
105	RARa	M413T	612	9694705
106	RARa	Q411X	610	1327285
107	RARa	R394W	589	9694705
108	RARa	M297L	475	9657734
109	RARa	L290V	442	9694705
110	RARa	R272Q	424	9657734
111	RARb	I403S fs*15	609	24075189
112	RARb	R387S/C	589	24075189

113	THRa	E403K/X	611	25670821
114	THRa	P398R/S	606	10022432 25670821
115	THRa	F397fs-406X	605	27144938
116	THRa	C392X	597	27144938
117	THRa	V390A	595	22507269
118	THRa	M388I	593	22507269
119	THRa	R384H	589	27144938
120	THRa	A382fs-388X	587	27144938
121	THRa	C380fs-387X	585	27144938
122	THRa	M369V	573	22507269
123	THRa	N359Y	563	26303090
124	THRa	Y352C	554	22507269
125	THRa	E350K	551	22507269
126	THRa	K337R	538	22507269
127	THRa	S305P	495	22507269
128	THRa	K288E/H	390	22507269
129	THRa	S271I	433	22507269
130	THRa	A264V	426	22507269
131	THRa	A263S	425	27144938
132	THRa	E245V	407	22507269
133	THRa	A225T/G	387	22507269
134	THRa	E213D	375	22507269
135	THRa	Q187X	349	22507269
136	THRa	H184Q	346	22507269
137	THRa	S183N	345	22507269
138	THRb	E460K	614	25905294
139	THRb	F459C/L	613	19268523 20237409
140	THRb	E457G	611	24722129
141	THRb	L456S	610	22507269
142	THRb	F455S	609	19299458
143	THRb	L454fs-463, L454V	608	17596672 8990194
144	THRb	P453H/A/L/T/S	607	2153155, 8040303, 19268523, 18561095
145	THRb	P452ins, P452R	606	8040303 24722129
146	THRb	F451S/I/L	605	27034829
147	THRb	P447T	598	19268523
148	THRb	C446R	597	22507269
149	THRb	M442V/T	593	19378427
150	THRb	L440P	591	19378427
151	THRb	R438H/C/P	589	8040303

				30027432
152	THRb	H435L/Q/Y/R	586	11889175 11701737
153	THRb	I431T/M	582	11889175 19268523
154	THRb	R429Q/W	580	8040303 12006711
155	THRb	D427G	577	22507269
156	THRb	T426I	576	10660344
157	THRb	F417L	567	22507269
158	THRb	H412R	562	22507269
159	THRb	K411E	561	22507269
160	THRb	F403L	550	22507269
161	THRb	A387P	534	27034829
162	THRb	S380F	516	22507269
163	THRb	M379T	515	22507269
164	THRb	L373P	509	22507269
165	THRb	S350L	486	24906004
166	THRb	V349M	485	18363280
167	THRb	V348E	484	8889584
168	THRb	G347E/A	483	1661299 17827792
169	THRb	L346F	482	19268523
170	THRb	G345R/V/S	481	25905294
171	THRb	G344E/A	478	19435825 21795843
172	THRb	K342I	476	15886199
173	THRb	L341P	475	19268523
174	THRb	Q340H	474	23806029
175	THRb	R338W/L	472	8514853 8040303
176	THRb	Del337T	445	1653889
177	THRb	A335P	443	19268523
178	THRb	E333D	441	17177139
179	THRb	G332R/E	440	8040303
180	THRb	N331D	439	19268523
181	THRb	L330S	438	10724359
182	THRb	T329I	437	19820907
183	THRb	T327A	435	19378427
184	THRb	Y321C/H	429	11756220 8040303
185	THRb	R320H/G	428	1314846 30027432
186	THRb	A318D	426	11889175
187	THRb	A317T/S	425	25738994 8889584

188	THRb	R316H/C	424	8381821 22319036
189	THRb	M313T/V	421	19268523 30027432
190	THRb	E299K	407	22507269
191	THRb	K289M	397	22507269
192	THRb	I280S	388	22319036
193	THRb	A279E	387	19378427
194	THRb	I276L	384	21795843
195	THRb	T273A	381	22507269
196	THRb	A268G	376	19268523
197	THRb	V264D	372	9092799
198	THRb	Q252R	360	11756220
199	THRb	I250T	358	19378427
200	THRb	R243W	351	9141558
201	THRb	Q235X	343	22507269
202	THRb	A234T	342	26273722
203	THRb	W219L	327	27034829
204	LXRa	R415Q	580	27253448
205	VDR	V346M	533	24246681
206	VDR	E329K	505	24246681
207	VDR	G319V	495	19169476
208	VDR	Q317X	493	12468277
209	VDR	I314S	490	8961271
210	VDR	H305Q	481	24246681
211	VDR	Y295X	445	24246681
212	VDR	W286R	436	24246681
213	VDR	R274L/H	424	24246681
214	VDR	I268T	418	24246681
215	VDR	L263R	413	24246681
216	VDR	Q259P/E	409	19169476 24246681
217	VDR	F251C	401	24246681
218	VDR	K246(3bp deletion)	396	24246681
219	VDR	L227P	377	24246681
220	VDR	R158C	311	24246681
221	VDR	Q152X	305	24246681
222	VDR	T146I	299	24246681
223	MR	L979P	646	12788847
224	MR	E972G	624	16954160
225	MR	I963fs994X	615	27725360
226	MR	A958fs1013X	610	16954160
227	MR	L924P	571	11134129
228	MR	R861X	498	19344080
229	MR	S818L	429	16954160

230	MR	S815R	426	16972228
231	MR	S810L	421	12538613
232	MR	S805P	416	16972228
233	MR	Q776R	387	12788847
234	MR	N770K	381	16972228
235	MR	L769P	380	16972228
236	MR	P759S	370	16972228
237	HNF4A	M364R	612	17407387
238	HNF4A	R303H	543	17407387
239	HNF4A	E276Q	515	10389854
240	HNF4A	D206Y	408	17407387
241	SF-1	L437Q	590	21078366
242	SF-1	D380Y	518	21078366
243	SF-1	V355M	493	21078366
244	SF-1	D293N	408	21078366
245	SF-1	W279X	394	21078366
246	RORa	R462Q	562	29656859
247	RORa	S409R	495	29656859
248	RORa	R340Pfs*17	397	29656859
249	RORa	Q315Lfs*51	373	29656859
250	RORb	Thr417del	580	27352968
251	RORg	Q441X	541	26160376
252	RORg	Q329X	389	26160376

Results

The phylogenetic analysis sorted NR LBDs into four distinct monophyletic branches: the steroid hormone receptor‐like cluster, the retinoid X‐like and steroidogenic factor‐like receptor cluster, the thyroid hormone‐like receptors cluster, and the nerve growth factor‐like/HNF4 receptor cluster **(Figure 6)**. The steroid hormone receptor-like cluster features all known SHRs and is divided into various subclusters. Unsurprisingly, GR, MR, PR, and AR LBDs were found to be closely related to ERs. As expected, ERs were separated in ERa and ERb, with ERa being surprisingly split into two groups. A more in-depth structural analysis was conducted in order to investigate this anomaly, with the results being exhibited later on in the text. Additionally, the SRH‐like branch of receptors displayed a close relationship with the second branch of the retinoid X‐like and steroidogenic factor‐like receptor cluster. This is the first time such a linkage has been observed. The second monophyletic branch consists of retinoid-x-receptors (RXRs), the liver receptor homolog-1 (LHR1), the steroidogenic factor 1 (SF1), and the ultraspiracle protein (USP) subunit of the ecdysone receptor. LHR1 and SF1 belong to the steroidogenic factor‐like subfamily of NRs and play a critical role in steroidogenesis (192). RXRs and their corresponding homolog in drosophila USP mainly participate in NR heterodimerization (193). The third monophyletic branch, which features thyroid hormone‐like receptors, includes thyroid hormone receptors (THRs), peroxisome proliferator-activated receptors (PPARs), liver X receptors (LXRs), the vitamin D

receptor (VDR), retinoic acid receptors (RARs), the farnesoid X receptor (FXR), RAR-related orphan receptors (RORs), the Rev-Erb alpha receptor (RevErb), and the EcR subunit of the ecdysone receptor. It should be noted that previous research has shown that the EcR subunit of the ecdysone receptor is closely related to FXR, based on the similarities found in their DBDs (194). It appears that the EcR/USP heterodimer is the arthropod analog of the FXR/RXR heterodimer. This specific analysis also highlighted that each subunit of the heterodimer belongs to a different monophyletic branch. Lastly, the fourth monophyletic branch was related to the thyroid hormone-like receptors monophyletic branch and included the hepatocyte nuclear factor 4 alpha (HNF4a) and Nur77 receptors. It should be noted that several GR structures were pretty different from other NRs.

Figure 6. A phylogenetic tree of NR LBDs. Four distinct monophyletic branches are observable: the steroid hormone receptor‐like branch (colored green), the retinoid X‐like and steroidogenic factor‐like receptor branch (colored red), the thyroid hormone‐like receptors branch (colored blue), and the nerve growth factor‐like/HNF4 receptor branch (colored yellow)

Seven signaling motifs were found during the study of the consensus sequences of NR LBDs **(Figure 7A)**. A similar analysis aimed specifically at SHRs also showcased

some important information regarding NR function **(Figure 7B)**. The first motif, termed motif A, occupies positions 378 to 385 in the sequence alignment and displays an LLxxL sequence. This sequence is an inverse NR-box (LxxLL). An NR-box, also known as an LxxLL-related motif, is a short motif found in multiple NR coactivators and allows direct interaction of coactivators with nuclear receptors (195). NR-boxes have been found on unique NRs, such as SHP (NR0B2), and appear to influence other NRs' functions (196). An inverse NR-box also seems to participate in NR activation, although rarely (197). According to NCBI's conserved domains database (CDD) (198), this specific region seems to be of great importance to NRs and is a main ligand interaction site. This motif is also present in SHRs. The second motif, termed motif B, occupies positions 391 to 401, and a query in CDD showcased that this region is critical for coactivators' function in SHRs. A PDB structure like 1L2I for ERa displays that this region is important in interacting with proteins featuring an NR-box. This motif's impact is also supported by the effect a mutation in this region may have on coactivators' function, with a prime example being GR V575G (198). This mutation seems to be located in GR's AF-2 surface, whose role is the attraction of a coactivator featuring an NR-box. Such a mutation hinders the interaction between NR and the corresponding coactivator. The third motif, termed motif C and displaying a sequence of LxxDDQ, occupies alignment positions 404 to 413, and also seems to impact coactivator function. Particularly, along with an R residue present in alignment position 402 it forms a structure specific to steroid hormone receptors GR, AR, and PR. This motif was first described in a paper by Bledsoe et al., which showcased that this motif takes part in the creation of the glucocorticoid receptor's second charge clamp. This structure is vital for specificity for the third TIF2 NR-box motif and seems to influence ligand binding and selectivity (199). The residues responsible for the second charge clamp appear to be missing from the remaining SHRs (ERs and MRs). The fourth motif, termed motif D, occupies alignment positions 512 to 516, with the region covered being a part of the highly conserved C-terminal end of the eighth helix structure in SHRs. This region seems to have a significant role in ligand binding since mutations like GR's L672P and AR's L813P result in the complete absence of ligand binding (200, 201). Additionally, it has been shown that resulting mutant proteins are potentially prone to a higher degradation rate (201). The fifth motif, termed motif E, occupies alignment positions 546 to 550, and, just like motif A, is an inverse NR-box. The sixth motif, termed motif F, occupies alignment positions 568 to 575 and is an NR-box (LxxLL) found in all SHRs, with the ERa LxxLL occupying positions 568 to 572. The seventh motif, termed motif G, occupies alignment positions 601 to 613. Specifically, it contains an ERa LxxLL motif on alignment positions 601 to 609, a PPAR LxxLL motif on alignment positions 605 to 609, and an LLxxL ERa motif on alignment positions 609 to 611. The existence of an LxxLLLxxL motif, i.e. a succession of an NR-box and its inverse, on ERa is quite an interesting observation. A mutation on motif G in ERa may lead to various pathological conditions in humans, with ERa Y537S and its association with breast cancer being a prime example (202).

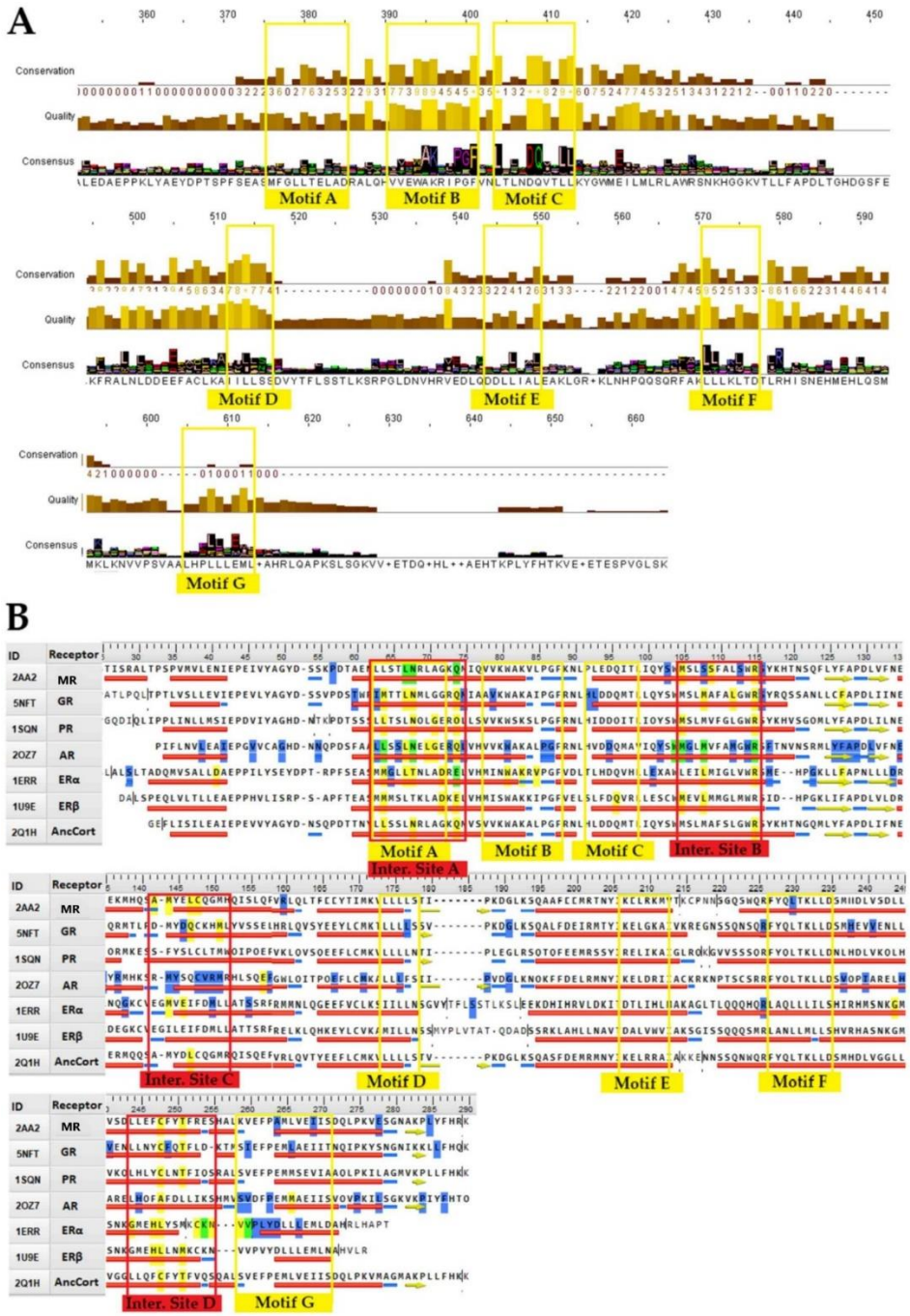

Figure 7. Conserved signaling motifs found on NR LBDs and conserved signaling motifs and interaction sites found on SHR LBDs A) Sequence alignment of all nuclear receptors' ligand-binding domains received from PDB. Parameters such as amino-acid quality and conservation are also present. The seven conserved signaling motifs are enclosed in yellow rectangles. B) Sequence alignment of steroid hormone receptors' ligand-binding domains with the conserved motifs found in all NRs being enclosed in yellow rectangles and the four interaction sites specific to SHRs being enclosed in red rectangles. Several amino-acid residues have also been highlighted to showcase distinct abilities. Specifically, yellow-colored residues are interaction points, blue-colored residues are prone to mutation, while green-colored residues are both interaction points and prone to mutation. The PDB ID of the representative sequences used for each SHR are 2AA2 for the mineralocorticoid receptor (MR), 5NFT for the glucocorticoid receptor (GR), 1SQN for the progesterone receptor (PR), 2OZ7 for the androgen receptor (AR), 1ERR for estrogen receptor alpha (ERa), 1U9E for estrogen receptor beta (ERb), and the 2Q1H for the ancestral corticoid receptor (AncCort). NR: Nuclear receptor; LBD: Ligand-Binding Domain; SHR: Steroid Hormone Receptor; PDB: Protein Databank

A study of the mutations that naturally emerge on NRs **(Figure 8)** along with the mutation rate of each specific alignment position **(Table 5)** provided several insights. Highly conserved regions showcased low to no mutations. This was expected since their evolutional conversation is a sign of their important role in protein function. Natural mutations on highly conserved regions may have deliberating effects and even prove to be lethal, thus resulting in no surviving phenotypes. The majority of mutations on steroid receptors that were examined led to hormone levels alteration, specifically the levels of the hormone that acts as a receptor ligand. Several times, though, phenotypes that would imply a specific mutation on the protein may not showcase the expected mutation on the protein product. Such phenotypes may emerge due to epigenetic alterations on NRs, mutations on NR cofactors, or mutations on non-coding regions that affect enhancer function (203).

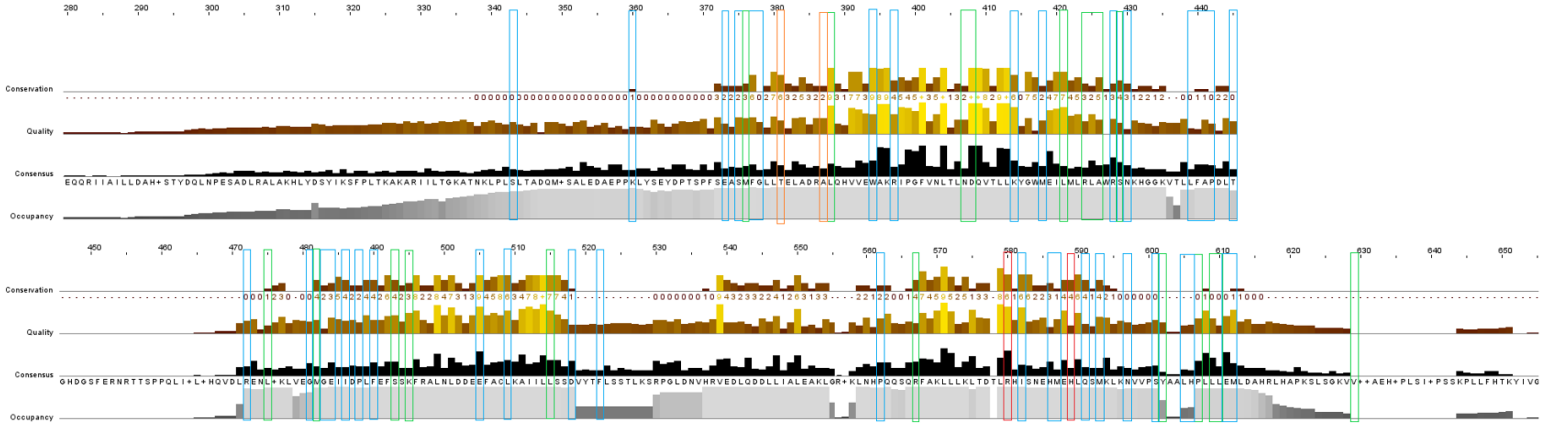

Figure 8. A schematic representation of the mutation rates of various sites on nuclear receptors. Specific sites on the multiple sequence alignment that are prone to mutations are enclosed into colored rectangles, with each color representing a different frequency (blue: mutations on two different NRs; green: mutations on three different NRs; orange: mutations on four different NRs; red: on five different NRs); NRs: nuclear receptors

Table 5. The mutation rate of each position on the nuclear receptor multiple alignment

Multiple sequence alignment position	**Frequency of Mutations (based on alignment position)**	**Receptor & Mutation**	**Pubmed ID of publications referencing the corresponding mutation**
624	3	AR L908F MR E972G HNF4 M364R	22334387 16954160 17407387
612	2	GR L753F RARa M413T	8316249 11050004
611	2	THRa E403K/X	25670821

		THRb E457G	24722129
610	3	RARa Q411X THRb L456S MR A958fs1013X	1327285 11756220 16954160
609	3	AR P839S RARb 1403 fs*15 THRb F455S	22334387 24075189 19299458
607	3	ERa D538G PPARg P467L THRb P453H/A/L/T/S	25838462 18713822 2153155, 8040303, 19268523, 18561095
606	2	THRb P452ins, P452R THRa P398R/S	8040303, 24722129 10022432, 25670821
605	2	THRa F397fs-406X THRb F451S/I/L	27144938 11756220, 24722129
602	3	GR I747M AR V890M ERa Y537N/C/S	12050230 22334387 25838462
601	2	AR S889del ERa L536P/R/Q	22334387 25838462 26183887
597	2	THRa C392X THRb C446R	27144938 8175986
593	2	THRa M388I THRb M442V/T	11756220 19378427
591	2	GR F737L THRb L440P	17635946 19378427
589	5	AR T878A RARa R394W RARb R387S/C THRa R384H THRb R438H/C/P	11906285 9694705 24075189 27144938 8040303, 30027432
587	2	AR H875Y THRa A382fs388X	22334387 27144938
586	2	PPARg H449L THRb H435L/Q/Y/R	26756202 11889175, 11701737
582	2	AR I870M THRb I431T/M	22334387 11889175, 19268523
580	5	GR H726R THRb R429Q/W LXRa R415Q RORb THR417del	26031419 8040303, 12006711 27253448 27352968
567	3	GR R714Q ERa R503W THRb F417L	20335448 24398047 11756220
562	2	THRb H412R	22507269

		RORa R462Q	29656859
522	2	GR G679S AR G821A	11589680 22334387
518	2	PPARg D396N SF-1 D380Y	17766367 21078366
515	3	GR L672P THRb M379T HNF4 E276Q	27120390 11889175 10389854
509	2	AR M808V/T/R THRb L373P	22334387 11889175
505	2	AR E804K VDR E329K	22334387 24818002
495	3	THRa S305P VDR G319V RORa S409R	11889175 19169476 29656859
493	3	ERa 432del-437X VDR Q317X SF-1 V355M	15583021 12468277 21078366
490	2	ERa M427I (+L429M) VDR I314S	15475371 8961271
488	2	AR R787X ERa M427I (+L429M)	22334387 15475371
486	2	AR C785Y THRb S530R	22334387 24217081
484	2	GR D641V THRb V348E	1704018 8889584
483	2	AR Y782D THRb G347E/A	22334387 1661299, 17827792
482	3	AR M781I PPARg F360L THRb L346F	22334387 25004973 19268523
481	2	THRb G345R/V/S VDR H305Q	25905294 24818002
475	3	AR R775H RARa M297L THRb L341P	22334387 9657734 19268523
472	2	ERa 411fs418X THRb R338W/L	16713253 8514853, 8040303
445	2	THRb del337T VDR Y295X	1653889 24818002
442	2	AR P767S RARa L290V	22334387 9694705
441	2	AR A766T THRb E333D	22334387 17177139
440	2	AR F765L THRb G332R/E	22334387 8040303

439	2	AR Y764C THRb N331D	22334387 19268523
430	2	AR F755L/V ERa M396V	22334387 16713253
429	3	GR 612 1bp del THRb Y321C/H MR S818L	20861124 11756220, 8040303 16954160
428	2	THRb R320H/G AR R753Q/X	1314846, 30027432 22334387
426	3	THRa A264V THRb A318D MR S815R	16434963 11889175 16972228
425	3	AR M750V THRa A263S THRb A317T/S	22334387 27144938 25738994, 8889584
424	3	RARa R272Q THRb R316H/C VDR R274L/H	9657734 8381821, 22319036 24818002
421	3	AR M764T THRb M313T/V MR S810L	22334387 19268523, 30027432 10884226
418	2	AR M743I/V VDR I268T	22334387 16059639
414	2	ERα E380Q PPARg K319X	25838462 10394368
408	3	AR D733Y HNF4 D206Y SF-1 D293N	22334387 17407387 21078366
407	3	PPARg 312fs315X THRa E245V THRb E299K	17011503 11889175 11756220
397	2	THRb K289M RORa R340Pfs*17	22507269 29656859
394	2	AR W719X SF-1 W279X	22334387 21078366
388	3	GR V571A AR L713F THRb I280S	11932321 22334387 22319036
387	4	ERa E353V THRa A225T/G THRb A279E MR Q776R	16713253 22507269 19378427 12788847
381	4	AR N706S PPARg Q286P THRb T273A MR N770K	22334387 10394368 11889175 16972228

378	2	AR S704C/G ERa 344insCyst	22334387 25838462
377	2	AR L702H VDR L227P	22334387 26422470
376	3	GR I559N AR L701M THRb A268G	8863343 22334387 19268523
375	2	PPARg R280P THRa E213D	29622583 11889175
373	2	GR T556I RORa Q315Lfs*51	21362280 29656859
360	2	AR V685I THRb Q252R	21362280 11756220
343	2	ERa S309F THRb Q235X	16713253 11889175

The above observations are also visible on the glucocorticoid receptor. No mutations were found on highly conserved regions, but some existing mutations did lead to debilitating effects on adrenocortical function. Specifically, several mutations on the GR LBD can cause Chrousos Syndrome, a genetic condition characterized by end-organ glucocorticoid insensitivity (204). Moreover, it is important to note that some mutations on GR LBD may have a dominant-negative effect, with mentioned mutations being more severe than other ones, since the effect normal protein function too.

Structure-wise, nuclear receptors are quite conserved. An in-depth analysis, though, highlighted two distinct canonical forms **(Figure 9)**. The first one appears to be more frequent in steroid hormone-like receptors, while the second one is more common in thyroid hormone-like receptors. The analysis also showcased that a subcluster that featured receptors USP, SF1 and LRH1 exhibit distinct structural features in regard to the two main canonical forms. Focusing on the SHR-like LBD canonical form, it is clear that it includes highly conserved structural domains, though estrogen receptors appear to be quite different from the rest of SHRs. It is also intriguing that ERb, particularly, is more similar to the rest of SHRs than ERa. Regarding structure specifics, all SHRs contain exactly four beta-strands, while the number of a-helixes is not consistent among all SHR PDB entries studied. It appears that a steroid hormone receptor may feature eleven or twelve a-helixes. The second major canonical form of the THR-like receptors LBD is also highly conserved, with a small number of differences amongst receptors. These differences lead to the formation of distinct subclasses, the PPAR-like, the ROR/THR, the VDR-like and the HNF4/Nur77-like. Regarding the effect of structural differences on NR's LBD function, a structural study focusing on GRs highlighted NRs' ability to form different activation states. The emergence of these different activation states is based on the position of the helix containing the AF-2 surface (199). Several factors may influence the activation state, including the bound ligand and the existence or lack of NR cofactors. Nevertheless, outliers did exist in the dataset, with three glucocorticoid receptor LBD structures (PDB ID: 3H52, 4LSJ, and 4MDD) and three LBD structures which belong to the NR2E

subfamily of nuclear receptors (PDB ID: 4LOG, 4XAJ, and 4XAI) distancing themselves from all other nuclear receptor ligand-binding domains. The GR LBD structures include the antagonist form of the receptor, while the three structures belonging to the NR2E subfamily of nuclear receptors feature a specific structural change in which the 12th helix occupies the coactivator binding site. A more thorough analysis of the GR structures showcased that they describe a specific antagonist form of GR in which its twelfth helix is dislocated and receptor function is disrupted (178).

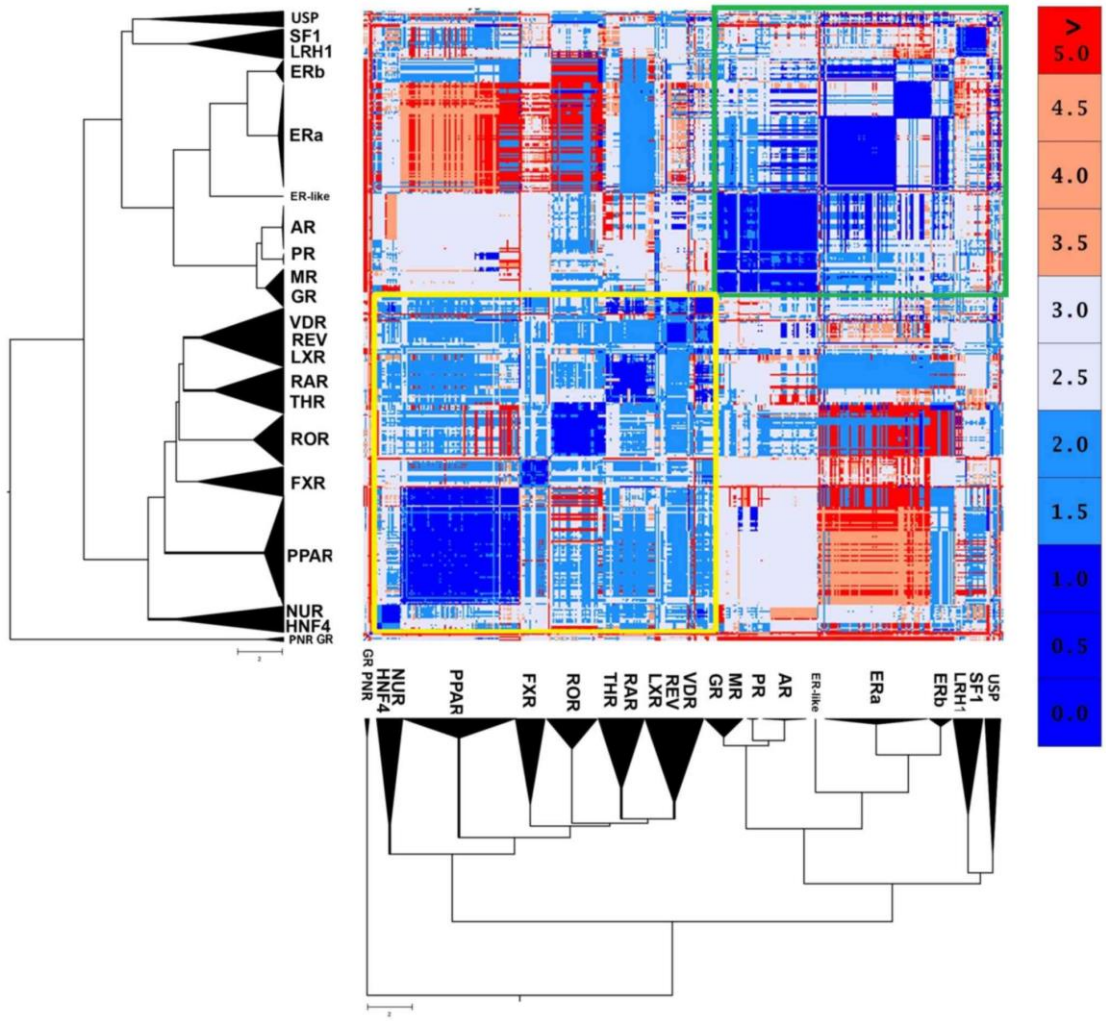

Figure 9. An in-depth structural analysis of nuclear receptors' ligand-binding domain. Structural similarity matrix of root mean squared deviation (RMSD) is displayed. The matrix separates NR LBDs into statistically significant clusters. The x and y-axis correspond to the structure order as described by the phylogenetic tree. Clusters colored blue indicate high structural similarity, while clusters colored red indicate low structural similarity. The yellow and green rectangle enclose the two major canonical forms. NR: Nuclear Receptor; LBD: Ligand-Binding Domain.

Nuclear receptors' ability to bind distinct ligands is, arguably, their more important functional characteristic (205). As mentioned -briefly- in the introductory chapter, NR ligands are small lipophilic, i.e., hydrophobic, molecules that bind their corresponding nuclear receptor's LBD hydrophobic pocket. Studying SHR ligands that were co-crystallized in the corresponding receptor's structure exhibit that the majority of ligands are also receptor-specific. The ability of MR to also bind

glucocorticoids has been already mentioned, but some other exceptions include mometasone furoate (MOF), which binds both GR and PR, plus R18, which binds both AR and PR. This is in accordance with the observation that receptors GR, MR, AR and PR are different from ERs and form their own subcluster in the phylogenetic tree. The PDB entry 1GS4 helps provide some interesting information regarding such similarities among SHRs (206). This specific entry describes an androgen receptor harboring mutations L701H and T877A. The T877A mutation causes the androgen receptor to bind specific anti-androgens but also progesterone and 17b-estradiol. Although this specific threonine seems to be unique to AR, based on SHR sequence alignment, the corresponding alignment position seems to have an important role in ligand interaction in all steroid receptors. Mutation L701H, on the other hand, causes AR to bind cortisol but severely impairs its ability to bind androgens. Based on SHR sequence alignment, this specific leucine is present in MR, PR, and AR, while the corresponding alignment position appears to be important for ligand interaction in GRs, ARs, PRs, and ERs (206).

Ligand analysis results showcased that there is a clear separation of ligands in three distinct clusters **(Figure 10)**. Those clusters are the MR/GR/AR/PR ligands cluster, the ER ligands cluster, and the USP/SF1/LRH1 ligands cluster. Two observations emerge during this separation. Firstly, the USP/SF1/LRH1 ligands cluster displays several similarities with the ER ligands cluster. Secondly, focusing on the ER ligands cluster, it seems that ERa is split into two different subclusters. These observations reinforce the view that ERs are quite distinct from MRs, GRs, Ars and PRs.

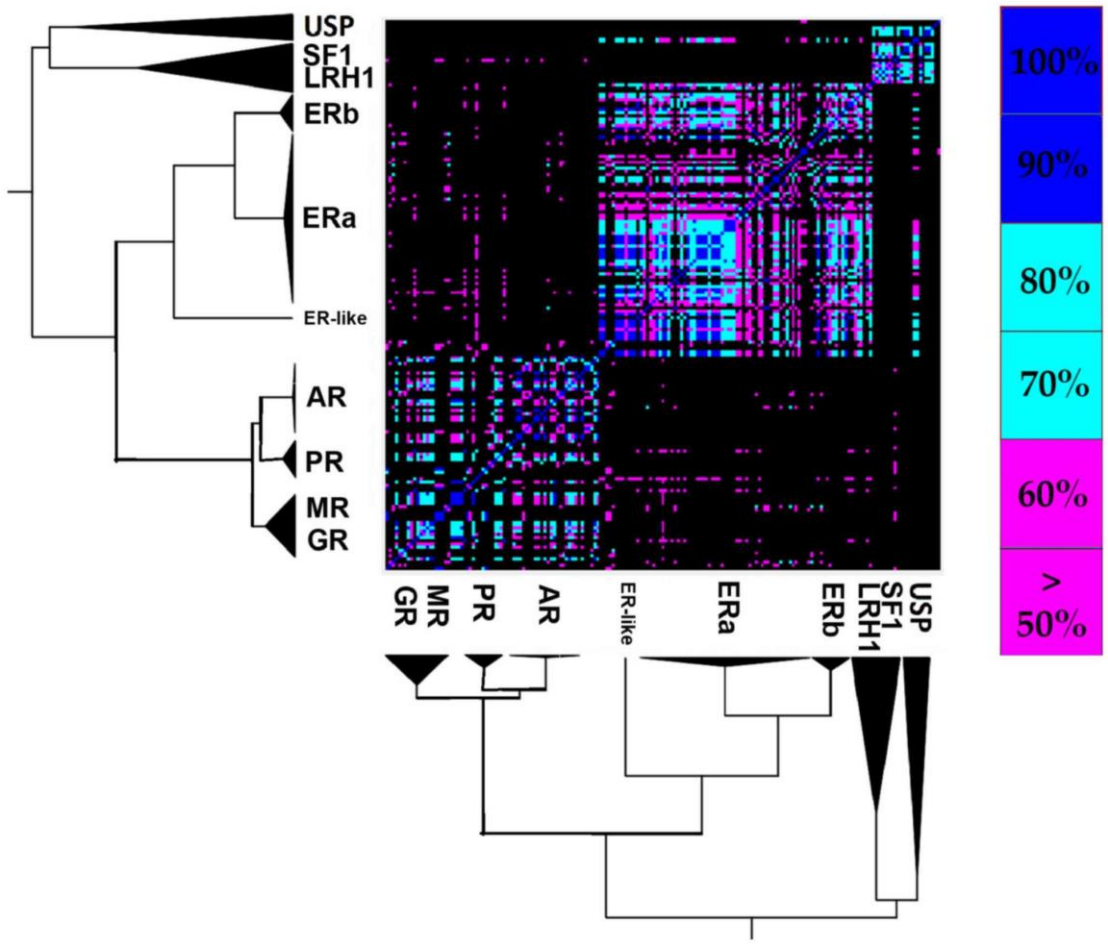

Figure 10. An analysis of NR ligands' chemical structure through the use of a chemical structure similarity matrix based on Tanimoto coefficient values. The Tanimoto coefficient varies between 0 when the fingerprints have zero bits in common and 1 when the fingerprints are identical. The corresponding ligands appear to split into three statistically significant different clusters. The x and y axes correspond to the ligands order, which is based on the order of the co-crystallized structures in the phylogenetic tree. Blue-colored clusters display strong structural similarity, while black-colored clusters display zero structural similarity. NR: Nuclear receptors

All of the above led to the implication that ERs and specifically ERa, displayed peculiar attributes, both on a sequence and on a structural level. The structure data implied that there was not a concrete group of type alpha estrogen receptors, but mentioned receptors were separated into two distinct subclusters **(Figure 11)**. This separation comes in contrast with the sequence analysis, which displayed high sequence similarity among type alpha estrogen receptors. A more in-depth look showcased that the first subcluster of ERa consisted of various mutant ERa entries and a small number of wild-type entries. Common mutations found in this subcluster were located on SHR alignment positions 381, 417, 530 and 536. The second subcluster was defined, mainly, by the Y537S mutation, with the ERa harboring mentioned mutant being referred to as ERa' from now on. The Y537S mutation is regularly found in breast cancer cells and has been implicated in resistance to a variety of endocrine therapies (207). This mutation is located in the helix harboring

the AF-2 surface of ERa' LBD and shifts the receptor equilibrium towards an agonist conformation regardless of the existence or not of a ligand. It is also important to note that ERa's structure seems to have a lot more in common with ERb (RMSD<2) than ERa, but on a sequence level ERa and ERa' display -as expected- minimal differences. This observation implies that the Y537S mutation leads to a conformational change that causes the receptor to attain a structure more similar to ERb. Specifically, Y537S leads to a 900 turn of the ERa helix containing the AF-2 surface. This new position of the helix is similar to ERb's AF-2 surface-containing helix position. This displacement seems to have an immense effect on receptor function, an expected outcome since the Y537S mutation is located in the highly important G motif as described by the sequence analysis conducted. Specifically, ligand analysis showcased that ERa' can interact with no specificity with all identified ER ligands, even ERb ligands. Several studies seem to agree with such findings (207, 208). Specifically, the study by Nettles et al., in which an ERα' (PDB: 2P15) was crystallized, concluded that this receptor could interact with a wider array of pharmacophores than previously thought. In summary, the Y537S mutation forces a conformational change in ERa LBD that leads to the emergence of a receptor (ERa') with the ability to bind ERb ligands, thus resulting in pathological conditions.

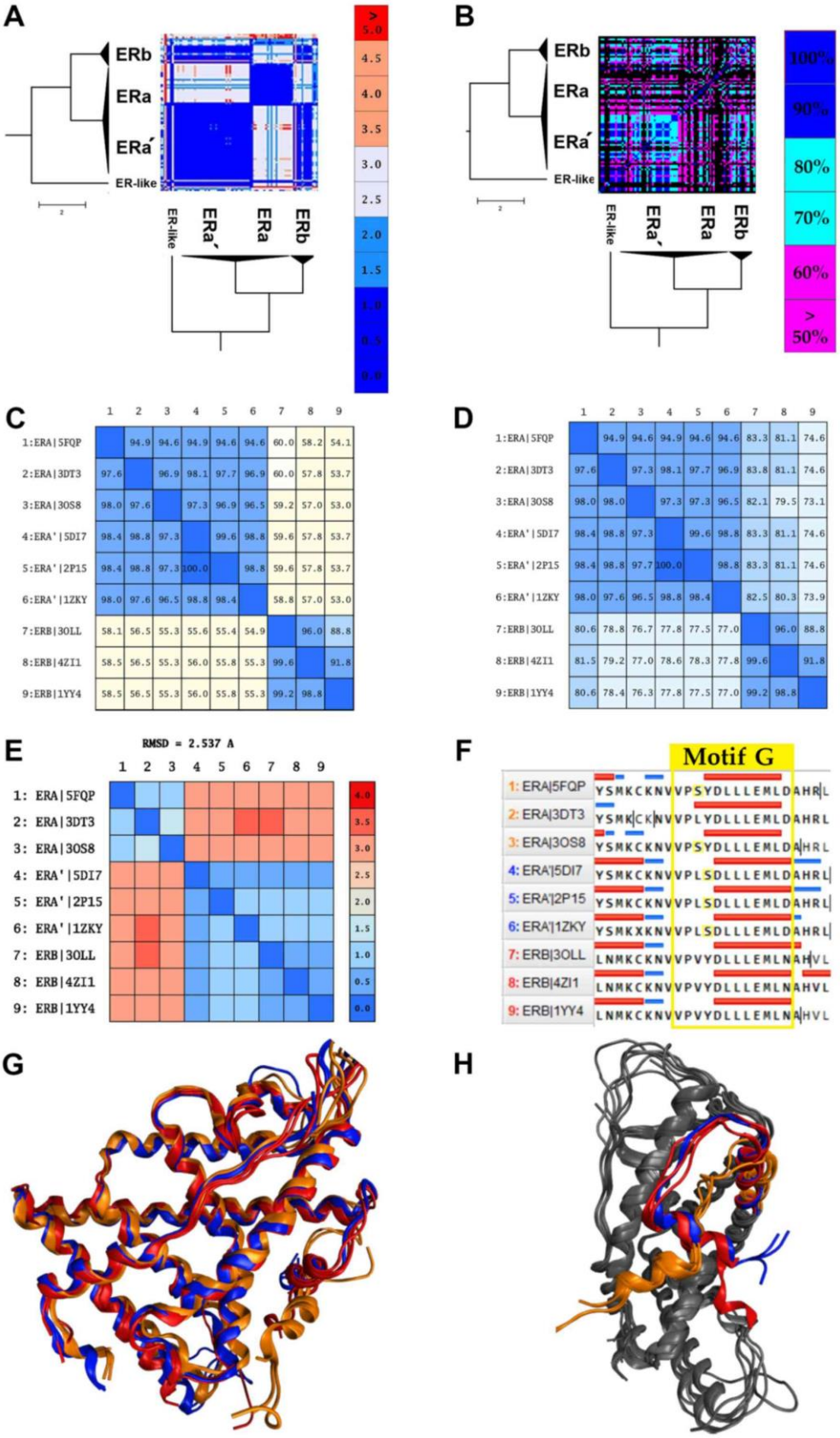

Figure 11. An in-depth structural and functional analysis of estrogen receptors (ERs). The representative structures for sections C to H are EFQP, 3DT3, and 3OS8 for ERa; 5DI7, 2P15, and 1ZKY for ERa'; 3OLL, 4ZI1, and 1YY4 for ERb. A) A structural similarity matrix of root mean squared deviation (RMSD) showed that ER LBDs are split into two statistically significant clusters, with estrogen receptor

alpha being further split into two subclusters, ERa and ERa'. B) An analysis of ER ligands' chemical structure through the use of a chemical structure similarity matrix based on Tanimoto coefficient values showcased that ERa' interacts with both ERa and ERb ligands. C) A sequence identity matrix based on nine representative sequences for ERa, ERa' and ERb. D) A sequence similarity matrix based on nine representative sequences for ERa, ERa' and ERb E) A structural similarity matrix of root mean squared deviation (RMSD) based on nine representative sequences for ERa, ERa' and ERb F) A multiple sequence alignment of the representative structures focusing on conserved signaling motif G, with known natural mutation points being colored yellow G) A ribbon presentation of estrogen receptors' representative structures superposed on each other, with ERa representatives being colored orange, ERα' representatives being colored blue, and ERb representatives being colored red H) A ribbon presentation of estrogen receptors' representative structures' AF-2 helix superposed on each other, with ERa representatives being colored orange, ERα' representatives being colored blue, and ERb representatives being colored red. LBD: Ligand-Binding Domain

Discussion

Since protein structure appears to be more conserved than protein sequence (209), hybrid phylogenetic analyses have gained popularity as methods of studying protein evolution. The study conducted made use of such an approach and did indeed highlight the evolutionary importance of structure, where small sequence alterations led to great structural with an immense effect on function. The current phylogenetic analysis showcased that NRs are separated into four distinct monophyletic branches based on their ligand-binding domain, the SHR-like cluster of receptors, the THR-like cluster of receptors, the RXR-like/SF1-like cluster of receptors and nerve-growth factor / HNF4-like cluster of receptors. Additionally, the ligand-binding domain of estrogen receptors seems to have distinct features compared to other SHRs, while GR, MR, AR and PR LBDs showcase major similarities.

Studying NR LBD sequences and identifying potential motifs is of utmost importance since conservation may indicate an important role in ligand-binding. An early study by Wurtz et al. found that a conserved signature sequence is vital in stabilizing the core of a canonical LBD (210). The current analysis found seven conserved motifs. The regions highlighted by the Wurtz et al. study coincide with some of the proposed motifs of the current phylogenetic analysis. The additional motifs proposed are mainly found on SHRs and are of the NR-box inverse and NR-box variety. The existence of such motifs on SHRs may initially seem odd, though the ability of some SHRS like GR to create homodimers or heterodimers may provide more context to such an observation. Since NR-boxes and their inverse can bind to specific receptor regions to modulate their function, there's a possibility for them to have a role in the interaction between SHR LBDs in order to influence each other's transcriptional capabilities or create a homo- or hetero- dimer. These specific sequences display moderate conservation in the alignment featuring all NRs. Lastly, LBD length doesn't deviate a lot among nuclear receptors.

Focusing on SHR LBDs adds to the phylogenetic analysis observation that ERs are quite unique since sequence analysis also showcases that they display distinct amino-acid sequences. Steroid hormone receptors exhibit four interaction sites in their LBDs, where sites A and B are found on all nuclear receptors in the current analysis and sites C and D are -mainly- found on SHRs. Finding ligand interaction

points and the effect of known point mutations on these sites may elucidate the functional properties of these regions. It was found that all ligand interaction points were prone to mutations. Although peculiar, it should be noted that interaction points are not necessarily an integral part of LBD structure maintenance. Moreover, these sites are more associated with NR selectivity and mutations in such regions may spearhead NR evolution. Sequence alignment also seems to support this theory since the interaction sites exhibit low sequence conservation. Lastly, it should be noted that the highly studied and well conserved A and B NR interaction sites are connected by conserved motifs A, B, and C.

The comparative analysis of SHR co-crystallized ligands showed that ligands are separated into three clusters, the SF1/LRH1/USP ligand-specific cluster, the GR/MR/AR/PR ligand-specific cluster, and the ER ligand-specific cluster. The main observation of this analysis is that the SF1/LRH1/USP ligand-specific cluster and the ER-ligand specific cluster showcase more similarities between them than each one of them with the GR/MR/AR/PR ligand-specific cluster separately. More importantly, ERa seems to split into two different sub-clusters, ERa and ERa'. ERa' appears to interact with zero specificity with all ligands co-crystallized with ERs, ERb ligands included. This ability may emerge due to a conformation change induced by the Y537S mutation in ERa, with mentioned mutation being heavily associated with breast cancer.

All the above information can prove useful for several real-life applications. The conserved motifs and interaction sites are intriguing drug targets. Researchers may develop, through in-silico approaches, molecules that can interact with those regions and alter receptor function. Sequence analysis highlighted the existence of NR-box and inverse NR-box motifs implicating a possible role in NR interaction and receptor homo- or hetero-dimerization. Mutation analysis showed the effects various mutations have on NR function, especially those found on conserved motifs.
Phylogenetic analysis provided novel information regarding NR LBD evolution. Last but not least, the structural study of both SHR LBDs and their corresponding ligands implied that the pathological effects of the Y537S mutation on ERa might be due to the structural change mentioned mutation promotes. Since this mutation has been heavily associated with cancer, this study's results may prove useful for elucidating breast cancer pathology and differences observed in drug response.

GR-related literature and GWAS Catalog data-mining and semantics study

Process and analysis of big data found in publicly available databases can help elucidate important information "hidden in plain sight". Data mining and semantics techniques can efficiently collect, annotate, and process such data (211). A wealth of information is hidden in literature databases like PubMed and curated collections of human genome-wide association studies like GWAS catalog (118, 212). A distinct pipeline was implemented to extract information regarding the glucocorticoid receptor and the regulators of its signaling from publicly available databases in an attempt to highlight understudied associations between mentioned molecular factors **(Figure 12)**.

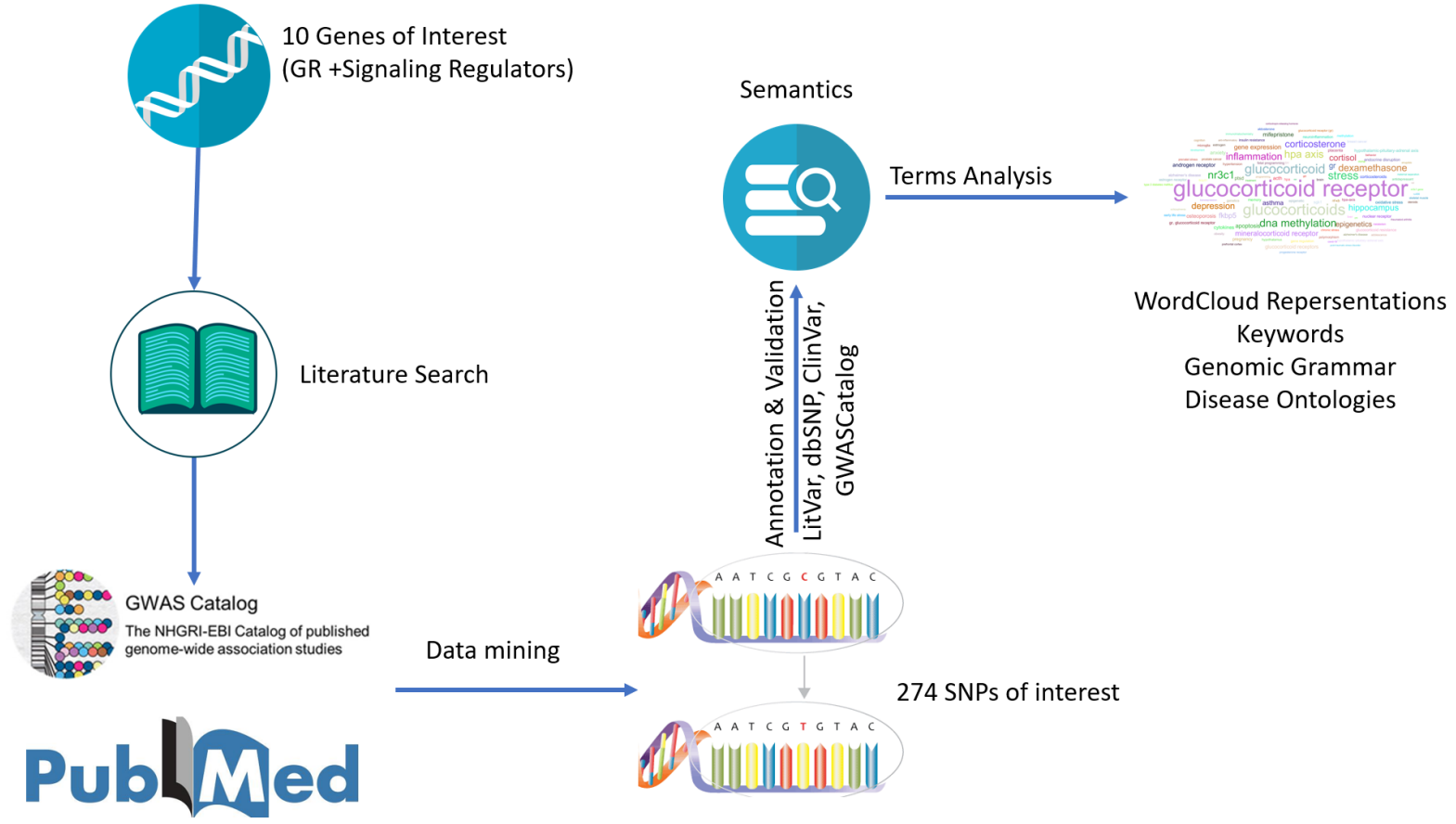

Figure 12 The pipeline followed during the current extraction of information regarding the glucocorticoid receptor and its signaling regulators

Methods of study

The glucocorticoid receptor, along with the main regulators of GR signaling **(Table 6)**, were used as the basis for a literature search. Specifically, the genes coding for the aforementioned factors were used as keywords in the PubMed database to receive relevant publications **(Table 7)**. Duplicate publications were removed after using a filtering algorithm and SNPs associated with the dataset studied were extracted. SNP terms were pinpointed and, consequently, any single nucleotide polymorphism that displayed a reference SNP ID number and was present in the dbSNP was obtained (213). GWAS Catalog was also used to receive additional SNPs that showcased an experimentally validated association with GR and its signaling regulators which were also present in the dbSNP database. These single nucleotide polymorphisms were named SNPs of interest.

Table 6 List displaying the main regulators of GR signaling along with the PMID of literature solidifying their role

Serial Number	Gene Name	PMID
1	FKBP5	19560279
2	FKBP4	32557257
3	HSP90(AA1)	28224564
4	PTGES3 (p23)	24345775
5	STIP1 (HOP)	32612187
6	HSP70	32612187
7	HSP40	30585227
8	NR3C2	28686058
9	BAG1	30585227

Table 7 A table displaying the keywords used to obtain information regarding the GR and its signaling regulators, along with the publications linked with each keyword and the cumulative publication for each factor

Serial Number	Gene Names	Literature	Total Literature
1	Glucocorticoid Receptor NR3C1	14.335 1.123	15.458
2	FKBP5 FKBP51 FKBP54 p54	831 403 5 702	1.941
3	FKBP4 FKBP52 p52	258 402 1.661	2.321
4	Heat Shock Protein 90	4.188	17.435

	HSP90 HSP90AA1	12.886 361	
5	PTGS3 Prostaglandin E Synthase 3 p23	101 15 2.598	2.714
6	STIP1 HOP STress-Induced Phosphoprotein 1 IEF-SSP-3521 p60 STI1	242 33.296 87 6 11.928 205	45.764
7	HSP70 HSPA4 APG-2 HS24/P52 HSPA4 HSP70RY HSPH2 HSPA1A HSPA1B	27.025 139 30 2 3 6 3 558 269	28.035
8	HSP40 DNAJB1 DnaJ Heat Shock Protein family member B1 HSPF1 Hdj1 Sis1	2.919 300 5 3 86 155	3.468
9	NR3C2 – Nuclear Receptor subfamily 3 group C member 2 MCR MLR MR – Mineralocorticoid Receptor	 304 5.274 7.589 8.893	22.060
10	BAG1 cochaperone 1 BAG-1 RAP46	747 1.078 14	1.839

The resulting SNPs of interest were annotated with information received from the LitVar database (214), ClinVar database (215), dbSNP and GWAS Catalog (118) and later validated. More specifically, the LitVar database was used to identify the most co-occurred entities regarding disease, chemical agents and SNPs in text featuring the SNPs of interest, the ClinVar database to elucidate if SNPs of interest displayed an association with a human pathological condition, the GWAS Catalog database to uncover if SNPs of interest were associated with a specific trait and, finally, the dbSNP to find each SNP's of interest position in the genome and particularly the gene itself.

Semantics and term analyses took place in order to uncover information regarding disease ontologies, along with pinpointing the most common keywords and genomic grammar present in the literature studied **(Figure 13)**. All results are presented in WordCloud representations.

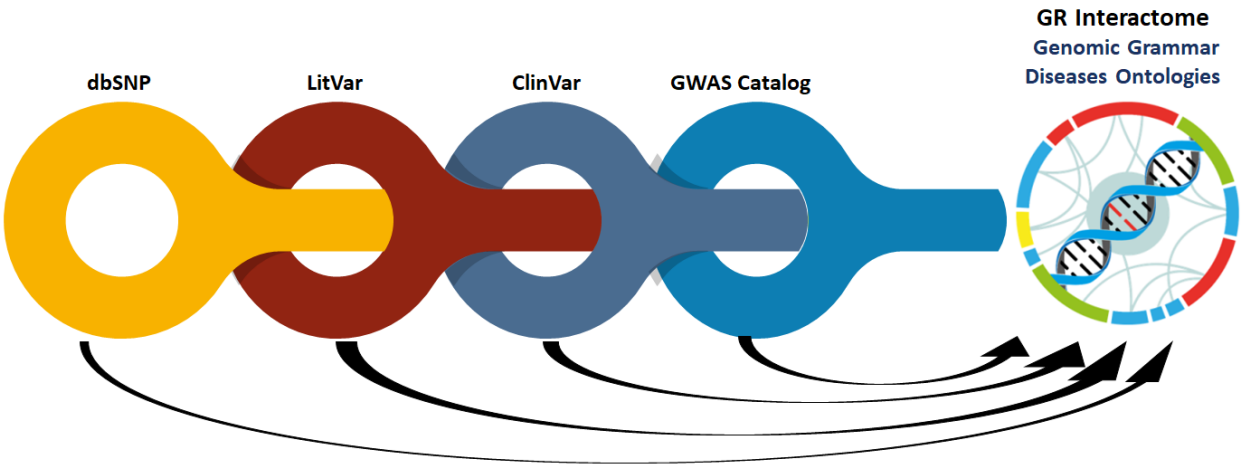

Figure 13 Information extraction from dbSNP, LtVar, ClinVar, and GWAS Catalog

Results

The methodology used resulted in 274 SNPs of interest, and the annotation process displayed an association with 247 diseases and 118 genes.

The SNPs present in the glucocorticoid receptor literature and GWAS Catalog were associated with several keywords **(Figure 14)**. Most of these keywords can be separated in distinct groups, specifically groups featuring: a) terms such as HPA axis, stress and chronic stress, which highlights the receptor's role in the stress response system ; b) terms such as DNA methylation and epigenetics, which highlights the receptor's role in gene regulation c) terms such as inflammation and NF-kB, which highlight the receptor's role in immune modulation; d) terms such as fetal programming, which highlight the receptor's role in organism development e) other steroid hormone receptors, with terms such as mineralocorticoid receptor, progesterone receptor and androgen receptor; f) terms highlighting the receptor's role in metabolism with terms such as insulin resistance and obesity; g) terms highlighting GR's role in neuropsychiatric disorders with terms such as depression, ptsd and schizophrenia; h) terms such as hippocampus, prefrontal cortex and microgloia, highlighting the receptor's role in brain architecture and neuroplasticity; i) members of the glucocorticoid receptor interactome, with terms such as FKBP5 and sgk1 and j) GR agonists and antagonists, with terms such as dexamethasone, and aldosterone. Apoptosis, i.e., programmed cell death (216),), is also present as a single word. Glucocorticoid signaling promotes, via the glucocorticoid receptor, pro-

apoptotic or anti-apoptotic processes depending on the cell type receiving the signal. A prime example is the effect GCs have on heart tissue, specifically endothelial cells and cardiomyocytes, with both cells displaying a critical role in the circulatory

system. Specifically, glucocorticoids seem to have a pro-apoptotic effect on endothelial cells and an anti-apoptotic effect on cardiomyocytes (217). Additional terms are also present, including various pathological conditions like Alzheimer's disease, asthma, prostate cancer, COVID-19, and osteoporosis, each having a different association with GCs or GR. Alzheimer's disease is, as previously mentioned, a neurodegenerative disease where glucocorticoids may play an important role. The mainstay of asthma therapy is synthetic glucocorticoids which act via the GR receptor (218). As mentioned before, the role glucocorticoids and GR play in cancer is complex and prostate cancer is no exception. GCs have been used in prostate cancer to slow disease progression, offset therapy side-effects and reduce pain (219). Nevertheless, some malignancies take advantage of GR signaling in order to promote resistance to anti-androgen therapies (220). COVID-19 is a severe infectious disease of the respiratory system with symptoms varying from mild and flu-like to lethal acute respiratory distress syndrome (ARDS) (221, 222). One of the main characteristics of COVID-19 is an excessive inflammatory response which may lead to an out-of-control host response resulting in the damage of respiratory cells (223). Glucocorticoids administration has been proposed in managing COVID-19 due to their anti-inflammatory abilities (222). Exposure to glucocorticoids may lead to fracture and bone loss. Particularly, GC-induced osteoporosis is the most common form of secondary osteoporosis (224).

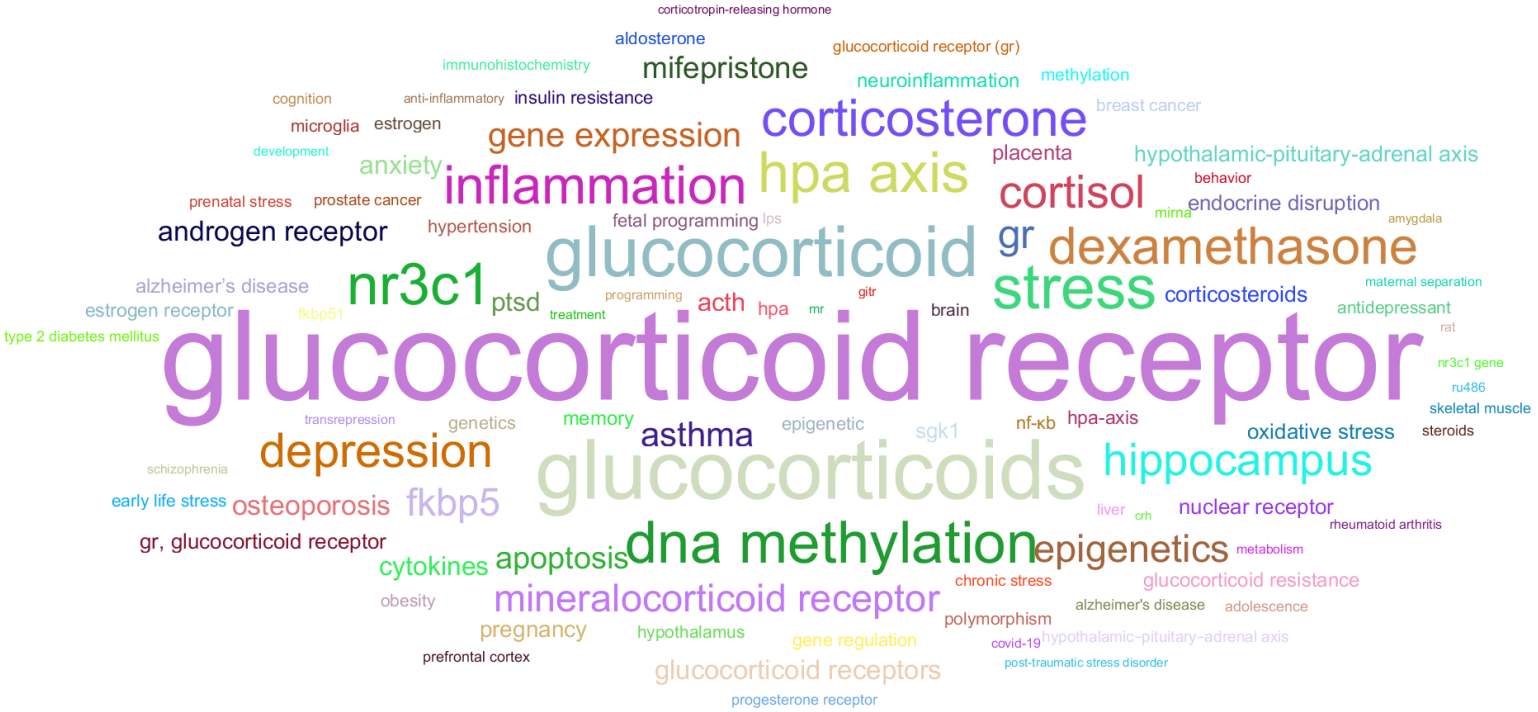

Figure 14 A WordCloud representation of the keywords found in GR literature

The SNPs found in GR literature and GWAS Catalog were studied in conjunction with various pathological conditions **(Figure 15)**. The pathological conditions most studied in conjunction with GR biology are - as expected – neuropsychiatric disorders, such as depression and post-traumatic stress disorder (PTSD), plus metabolic disorders like diabetes mellitus and obesity. Pathologies such as systemic erythematosus

lupus, asthma and rheumatoid arthritis are also associated with GR study, which may be due to the fact that the administration of synthetic glucocorticoids is one of the main treatments for such disorders (225). Glucocorticoid research seems to also overlap with studies on cardiovascular diseases, such as hypertension, possibly due to the aforementioned influence of the stress response system on cardiovascular function. Somewhat unexpected though, is the immense study of various neoplasms or wounds and injuries in conjunction with GR. The role of glucocorticoids in cancer pathology and pathogenesis is, as mentioned, still under evaluation. Although the glucocorticoid receptor is not considered an oncogene, glucocorticoids administration has been shown to arrest growth and induce apoptosis in lymphoid tissue via receptor signaling in certain patients (226). The increasing interest in uncovering the mechanisms which regulate the glucocorticoids' effect on cancer may be the reason why a great part of GR literature focuses on the receptor's role in cancer. Wounds and injuries, on the other hand, are heavily associated with inflammation, since inflammation is the first step towards wound healing (227). The glucocorticoid receptor is known to have a major role in inflammation and thus make take part in the regulation of the healing process.

Figure 15 A WordCloud representation of diseases found in GR literature

SNPs found in GR signaling regulators were studied in conjunction with specific genes in literature and GWAS Catalog **(Figure 16)**. Most of the resulting genes are regulators themselves (FKBP5, HSPA1L, STIP1). Additional genes included code for regulators of the HPA axis, such as CRHR1; immune system regulators, such as CFH and NFKB2; factors influencing brain architecture, such as BDNF-AS and NTRK2; factors that influence metabolism, such as APOE and FTO; and the mineralocorticoid receptor. Genetic locations responsible for non-coding RNAs were also observed.

These locations were loc112267956, loc101929309, Loc105378525, and miR4761. Non-coding RNAs are known to play a major role in gene regulation (228) and thus may influence gene expression of GR signaling regulators or assist in GR's action as a transcription factor (229). Highly prominent genes in the resulting dataset were also VEGFA and POLR1C. The VEGFA gene is responsible for the vascular endothelial growth factor (VEGF), which has an essential role in both physiological and pathological angiogenesis (230), while POLR1C codes for a subunit of RNA polymerases I and III (231). GCs are known to have an angiostatic effect and glucocorticoids treatment has been shown to influence VEGF mRNA levels (232). RNA polymerases I and III are mainly known to participate in ribosomal RNA transcription (231). Several studies from the 1980s had highlighted that glucocorticoids influence rRNA synthesis in animal models, though results were conflicting on whether GC administration led to an increase or decrease of rRNA synthesis (233, 234).

Figure 16 A WordCloud representation of genes associated with GR signaling regulators in literature

Lastly, SNPs found in GR signaling regulators have been studied for their role in several diseases **(Figure 17)**. These diseases almost completely overlap with the diseases studied in GR literature. Several diseases are unique to SNPs found in GR signaling regulators, such as type 2 glycogen storage disease and non-alcoholic fatty liver disease. Type 2 glycogen storage disease, also known as Pompe disease, is a rare neuromuscular disease due to acid alfa-glucosidase (GAA) deficiency (235).). This disease is quite heterogenous and is classified into three forms, infantile,

childhood/juvenile and adult (236). The infantile form displays the most severe clinical manifestations, particularly skeletal and cardiac myopathy, which, unless treated, leads to early death, while childhood/juvenile and adult phenotypes display progressive myopathy with next to zero risk of hypertrophic cardiomyopathy (235). Enzyme replacement therapy is the most common approach to this disease, though several patients develop infusion reactions. Corticosteroids and antihistamines are administered to such patients, which may explain the association present in GR signaling with Pompe disease in the literature (237). Non-alcoholic fatty liver disease is an umbrella term that encompasses from simple steatosis to more progressive steatosis with associated hepatitis, fibrosis, cirrhosis and, in more severe cases, hepatocellular carcinoma (238). This disease is considered the hepatic manifestation of metabolic disorders and obesity. Glucocorticoids play a crucial role in non-alcoholic fatty liver disease since they seem to drive both pathogenesis and pathology (239). Another intriguing observation is that neoplasms studies are more present in GR signaling regulators SNPs than in GR SNPs, implying that the glucocorticoid receptor may play a more complicated role in cancer via indirect regulation. Other disease-related studies which are unique to GR signaling regulators research include Parkinson's and epilepsy, highlighting the role of GR signaling in proper brain function and polycystic ovary syndrome (PCOS). The inclusion of PCOS may be due to the effect glucocorticoids have on the hypothalamic-pituitary-gonadal axis, whose products have a key role in this disease's pathophysiology (13).

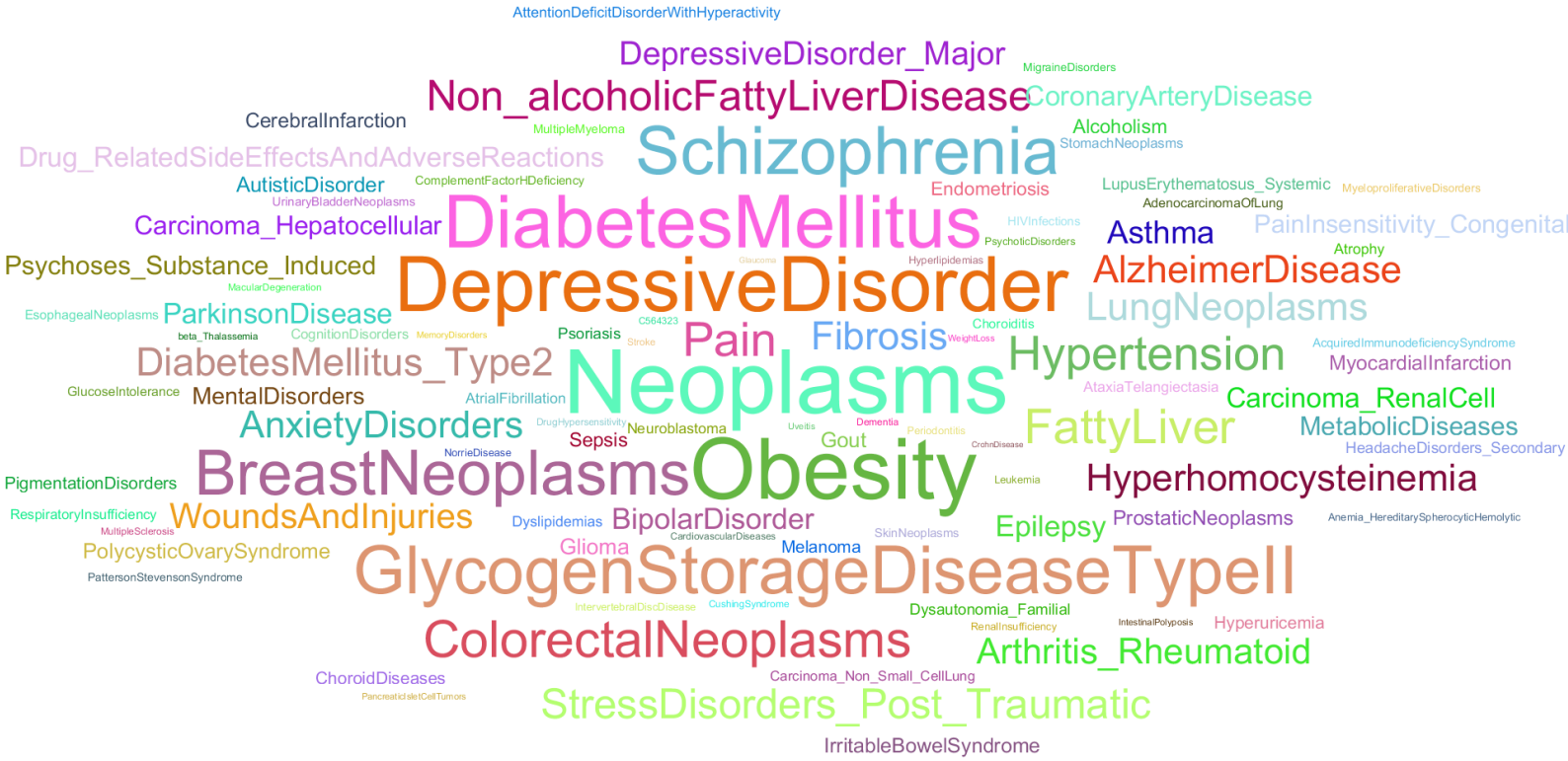

Figure 17 A WordCloud representation of diseases studied in conjunction with GR signaling regulators

Discussion

Studying GR and GR signaling mediators' GWAS data and literature, cements the importance of the glucocorticoid receptor and the stress response system in numerous physiological and pathophysiological mechanisms, with additional research implicating a possible role in cancer. From immune response and metabolism to stress response and proper brain function, GR seems to mediate some of the most important biological pathways. The diseases associated with GR or its signaling regulators are, as expected, associated with abnormalities on the aforementioned pathways **(Figure 15 and Figure 17)**. Moreover, the common use of synthetic glucocorticoids as anti-inflammatory drugs means that GR signaling is studied in conjunction with a wide variety of diseases. Some additional observations include the presence of non-coding RNAs such as loc112267956 and loc101929309 and the gene POLR1C in GR regulators' literature. These non-coding RNAs are produced by intronic regions of the FKBP5 gene. Although introns were thought to have no significant biological function, modern studies highlight their importance in various processes, from alternate splicing to gene regulation (240), with some research indicating that the miRNAs produced by introns may participate in negative self-regulation of gene expression (241). GCs' effect on the POLR1C, on the other hand, is quite intriguing, since although research on GR's influence on rRNA synthesis seems to stall in the 1980s, this gene is - in some way or another- heavily present in current GR signaling regulators literature. It is possible that heat shock proteins that play an essential role in both GR signaling and rRNA synthesis in response to heat shock may explain such an observation (242). Another possibility is that, since stress is thought to influence rRNA synthesis (243), the glucocorticoid receptor may in some way influence RNA polymerase I or III function. There is a possibility that RNA polymerase I is a downstream effector of GR signaling (244), since GR influences TBP-associated factor RNA polymerase I subunit A (TIF-IA). TIF-IA is a general transcription factor that connects other general transcription factors, such as the upstream binding factor (UBF) and selective factor 1 (SL1) ,with RNA polymerase 1 to initiate pre-ribosomal RNA transcription (245). Glucocorticoids are known to inhibit the c-Jun N-terminal kinase (JNK) pathway via GR signaling (246). JNK2 has the ability to induce the phosphorylation of TIF-IA (specifically on the threonine 200 residue), which in turn promotes RNA polymerase I function and rRNA synthesis (247). On the other hand, the glucocorticoid receptor is known to -mainly- inhibit mammalian target of rapamycin (mTOR) signaling (248). Inhibition of mTOR signaling is thought to inactivate TIF-IA and thus hinder RNA polymerase I function and rRNA synthesis (249, 250) **(Figure 18)**. These two pathways appear to lead to opposite results, therefore further research is required to identify the exact way glucocorticoids seem to play an important role in RNA polymerase I function.

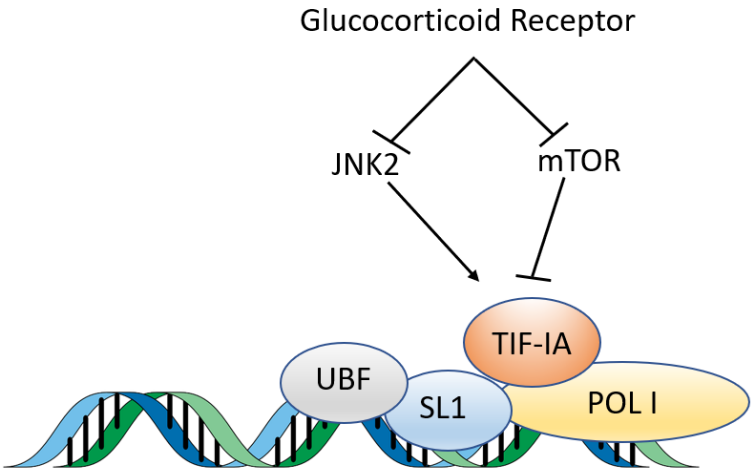

Figure 18 Potential GR actions that may influence RNA polymerase I function. GR may inhibit JNK2 function, and thus TIF-IA action. GR may also inhibit mTOR signaling and thus lead to TIF-IA activation (JNK2: c-Jun N-terminal kinase 2; mTOR: mammalian target of rapamycin ; TF-IA: TBP-associated factor RNA polymerase I subunit A; UBF: upstream binding factor; SL1: selective factor 1; POL I: RNA polymerase I)

Large population examination and extracting useful information

Evolutionary biology attempts to discern the way in which natural selection and other factors, such as random drift and mutations drive evolution (251). Population genetics is a scientific field whose goal is the elucidation of a population's genetic composition and the forces that underlie and alter mentioned composition (252). Thus, population genetics and evolutionary biology have overlapping goals, especially regarding the forces that promote phenotypic variation and, on a larger scale, evolution. Consequently, population genetic studies can provide useful information regarding the evolution of complex traits. The rapid advancements in mass sequencing technology have allowed population geneticists to systematically characterize SNPs, therefore aiding in their study of a population's genetic composition (253). Associating such SNPs with the distinct phenotypic characteristics of a population can also help pinpoint how specific genetic variations influence complex mechanism functions. Population genetics studies on the stress response system and the peculiarities each population showcases in its function may provide some important information regarding the interplay of various genetic networks that underlie the maintenance of homeostasis. Focusing on SNPs found in the GR interactome can help provide insights into the complex interplay of nuclear receptors transcriptional networks and their contribution to the maintenance of homeostasis since the glucocorticoid receptor is known to interact with other nuclear receptors and has a major role in the maintenance of homeostasis. A detailed pipeline was implemented in order to extract information in a precise and

efficient way from a dataset of Japanese individuals. The results produced were later used to compare Japanese and Korean populations **(Figure 19)**.

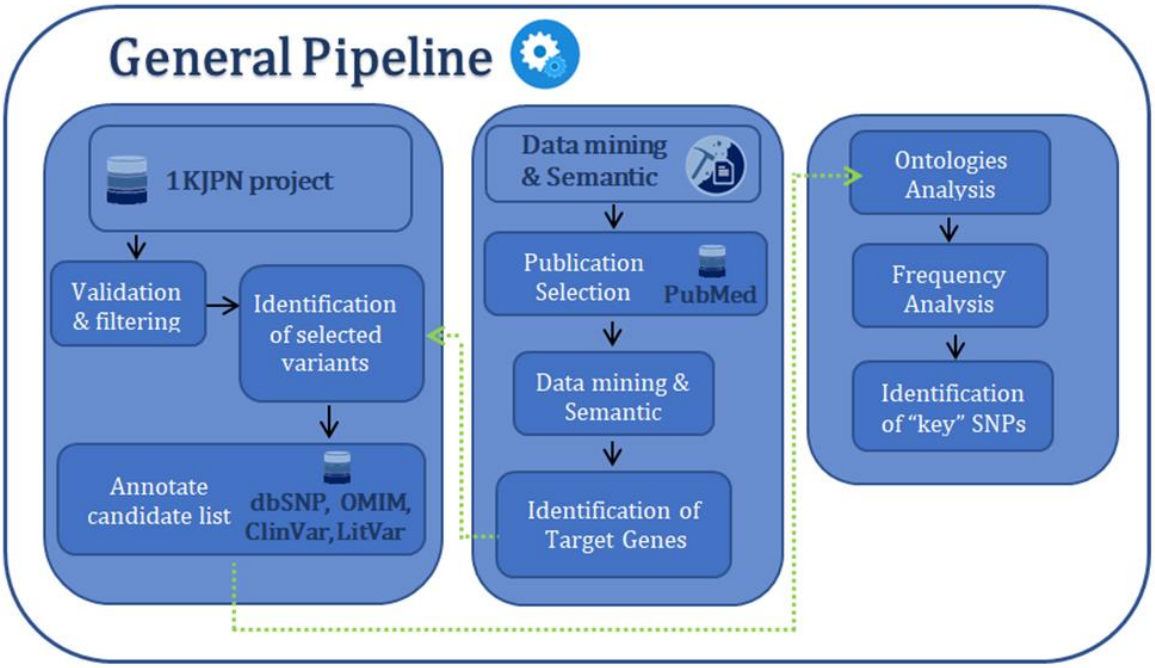

Figure 19 The pipeline followed during the current population comparison

Methods of study

The dataset used was a continuation of the 1KJPN project (254), specifically the 2017 update, which had reached a sample size of 3,554 (3,5K) Japanese individuals. The dataset received had already undergone a filtering procedure, with the SNPs used having 'passed' every filtering step **(Table 8)**. These SNPs were present on autosomes. Therefore factors that were located in sex chromosomes were not present. The identified single nucleotide polymorphisms included reference SNP ID number (rs ID) based on dbSNP if such rs ID existed (213). The genomic position of each SNP was based on the GRCh37/hg19 assembly.

Table 8 Filtering steps undertaken in the 3,5K Japanese dataset

Category	Total SNVs	Matched SNVs	Description
Step 1 (Multi-allelic)	50.099.977	165.439	Multi-allelic SNVs in 3.5KJPN but biallelic in 1KJPN and 2KJPN
Step 2	49.934.538	1.373.119	Depth filter (in a naive call, an alternative variant is detected but

			disappears with the sequence depth filter, e.g., miscall with CNV, somatic call, or misalignment)
Step 3	48.561.419	2.835.609	Depth filter (more than 10% of individuals do not fit into the reliable sequence depth range)
Step 4	45.725.810	6.969.032	SNVs in highly repetitive regions
Step 5	38.756.778	1.267.757	SNVs that are not detected in other alignment tools and variant callers
Step 6	37.489.021	421.306	The SNVs's hardy Weinberg equilibrium is less than or equal to 0.00001
Step 7	37.067.715	13.032.262	

A literature review was conducted and a comprehensive list that features 149 autosomal genes with an essential role in GR function or are prime examples of GR target genes, was composed **(Table 9)**. These genes contain, among others, nuclear receptors, molecular epigenetic regulators, GR cofactors and several enzymes. The genomic location of each gene was described based on the GRCh37/hg19 assembly.

Table 9. List of autosomal genes with an essential role in glucocorticoid receptor function

NR3C1	Glucocorticoid Receptor		5 NC_000005.9 (142657496..143113322, complement)
Factor	**Influence**	**Pubmed ID**	**Chromosome n Position**
BAG1	Interacts with hsp70, binds to hinge region, inhibits DNA binding and transactivation	9603979 11101523	9 NC_000009.11 (33252469..33264759, complement)
CDKs (1,2,5)	Different effects on GRE-containing promoters, based on phosphorylation sites (Serine 203 and Serine 226 decreased activation, Serine 211 enhanced activation)	19787703	10 NC_000010.10 (62538089..62554610) 12 NC_000012.11 (56360556..56366573) 7 NC_000007.13

			(150750899..150755052, complement)
TAT	GR is essential for TAT gene induction.	11420718	16 NC_000016.9 (71600754..71610998, complement)
GSK-3b	Leads to a conformation change in GR, attenuates GC signaling	19787703	3 NC_000003.11 (119540800..119813264, complement)
ERK2	Decreases receptor activity	9199329	22 NC_000022.10 (22113946..22221970, complement)
p38 (MAPK11-14)	Enhances GRE-related activity	15817653	22 NC_000022.10 (50702142..50708779, complement) 22 NC_000022.10 (50691331..50700248, complement) 6 NC_000006.11 (36098261..36112301) 6 NC_000006.11 (35995412..36079013)
MAPK8(JNK1)	Decreases receptor activity	12351702	10 NC_000010.10 (49514682..49647403)
Ubc9	Sumoylation/ Increases GR activity	12144530	16 NC_000016.9 (1357420..1377019)
LCK	Unliganded GR is part of a TcR-linked multiprotein complex containing Hsp90, LCK, and FYN	27169854	1 NC_000001.10 (32716840..32751766)
FYN	Unliganded GR is part of a TcR-linked multiprotein complex containing Hsp90, LCK, and FYN	27169854	6 NC_000006.11 (111981535..112194655, complement)
SUMO1	Sumoylation/Increased GR activity	12144530	2 NC_000002.11 (203070903..203103322, complement)
CHIP(STUB1)	Receptor downregulation and decreased transactivation	15761032	16 NC_000016.9 (730115..732768)
Mdm2	Takes part in GR degradation	12897156	12 NC_000012.11 (69201952..69239324)
DNA-PKcs	Phosphorylation of GR hinge region	9038175	8 NC_000008.10 (48685669..48872743, complement)
CEBPB	GR potentiates the action of CEBPB, Along with p21 it can inhibit cdk2 action	9817600	20 NC_000020.10 (48807120..48809227)

HDAC2	Influences GC sensitivity (overexpression leads to increased sensitivity)	23953592	6 NC_000006.11 (114257320..114292359, complement)
SGK1	GR upregulates its specific gene	23953592	6 NC_000006.11 (134490384..134639196, complement)
ZFP36	GR upegulates its specific gene	23953592	19 NC_000019.9 (39897487..39900052)
DUSP1	GR upegulates its specific gene	23953592	5 NC_000005.9 (172195093..172198203, complement)
β-arrestin (1,2)	GR regulates their gene expression (β-arrestin 1 +, β-arrestin 2 -)	23953592	11 NC_000011.9 (74971166..75062875, complement) 17 NC_000017.10 (4613789..4624795)
BGLAP	GR downregulates its specific gene	23953592	1 NC_000001.10 (156211951..156213123)
TBP	GR's AF-1 domain binds TBP, overexpression of TBP leads to stimulated expression of GR-driven reporters	16469772	6 NC_000006.11 (170863384..170881958)
CBP	Coactivator, Interacts with GR and p300 to form docking platform for transcription factors	19818358	16 NC_000016.9 (3775055..3930121, complement)
p300	Coactivator, Interacts with GR and CBP to form docking platform for transcription factors	19818358	22 NC_000022.10 (41488614..41576081)
Pcaf	Interacts with p300CBP and STAT3 to stimulate GR activity	27169854	3 NC_000003.11 (20081524..20195896)
NCOAs(1,2,3)	Coactivators that assist DNA expression, upregulation	19805480	2 NC_000002.11 (24714919..24993571) 8 NC_000008.10 (71021997..71316062, complement) 20 NC_000020.10 (46130601..46285621)
SMAD6	SMAD6 suppresses GR function	27169854	15 NC_000015.9 (66994674..67074338)
DAP3	Binds HSP90,increases transactivation activity	10903152	1 NC_000001.10 (155657693..155708801)
DAXX	Suppresses GR expression	12595526	6 NC_000006.11 (33286335..33290793, complement)
PP1 (PPP1CA, PPP1CB,	May reverse GR phosphorylation	19818358	11 NC_000011.9 (67165652..67169376,

PPP1CC)			complement) 2 NC_000002.11 (28974614..29025806) 12 NC_000012.11 (111157613..111180783, complement)
PP2(PPP2CA, PPP2CB)	May reverse GR phosphorylation	19818358	5 NC_000005.9 (133532148..133561950, complement) 8 NC_000008.10 (30643126..30670352, complement)
MED1	Enhances GR expression	10508170	17 NC_000017.10 (37560538..37607527, complement)
HNRNPU	Overexpression of HNRPU leads to GR inactivation	9353307	1 NC_000001.10 (245013602..245027827, complement)
HSP90 (HSP90AA1, HSP90AA2P)	Essential chaperone for GR function	28224564	14 NC_000014.8 (102547075..102606086, complement) 11 NC_000011.9 (27909718..27912639, complement) pseudogene
HSP70 (HSPA1A, HSPA1B, HSPA1L)	Essential chaperone for GR function	24949977	6 NC_000006.11 (31782952..31785719) 6 NC_000006.11 (31789964..31798032) 6 NC_000006.11 (31777396..31790093, complement)
HSP40 (DNAJA1, DNAJA2, DNAJA3, DNAJB1)	Increases the efficiency of the GR/chaperons complex	24345775 24949977 20453930 26245905	9 NC_000009.11 (33025209..33039905) 16 NC_000016.9 (46989274..47007625, complement)

			16 NC_000016.9 (4475806..4506776) 19 NC_000019.9 (14625581..14640087, complement)
HOP	Increases the efficiency of the GR/chaperons complex	10764743 24949977	11 NC_000011.9 (63952206..63972020)
p23	Increases the efficiency of the GR/chaperons complex	24345775 24949977	12 NC_000012.11 (57057125..57082138, complement)
MR	Heterodimerization with GR and coordinates transcription	11154266	4 NC_000004.11 (148999915..149365850, complement)
Cytochrome p450 enzymes (CYP3A4, CYP3A5, CYP2C8, CYP2C9, CYP2C19)	GR regulates the enzymes' expression	24451000	7 NC_000007.13 (99354583..99381811, complement) 7 NC_000007.13 (99245813..99277636, complement) 10 NC_000010.10 (96796529..96829255, complement) 10 NC_000010.10 (96698350..96749486) 10 NC_000010.10 (96522463..96612671)
P-glycoprotein	GR regulates its expression	24451000	7 NC_000007.13 (87133179..87342639, complement)
FKBP4(FKBP52)	Regulates GR signaling, possibly positive regulation	19818358	12 NC_000012.11 (2904108..2914589)
FKBP5(FKBP51)	Regulates GR signaling, possibly negative regulation	19818358	6 NC_000006.11 (35541362..35696360, complement)
NRIP1	Negatively regulates the activity of GR	12773562	21 NC_000021.8 (16333556..16438224, complement)
CLOCK	Represses GR-induced transcriptional activity	19818358	4 NC_000004.11 (56294068..56413076, complement)
BMAL1	Represses GR-induced transcriptional activity	19818358	11 NC_000011.9 (13299325..13408813)

AP-1 (specifically c-Fos)	GR weakly interacts and inhibits AP-1 dependent transcription. Specifically, GR binds cFos/cJun via a sequence unique to cFos	27169854	14 NC_000014.8 (75745477..75748937)
NF-κB	GR interacts with NF-κB through the second zinc finger of the ligand-binding domain and acts negatively on the p65/RelA subunit of NFκB.	19818358	11 NC_000011.9 (65421067..65430443, complement)
POU2F1	GR interacts with POU2F1 in order to bind to distal nGRE	9891005	1 NC_000001.10 (167190066..167396582)
POU2F2	GR interacts with POU2F2 in order to promote the binding of POU2F2 o specific sequences	10480874	19 NC_000019.9 (42590262..42636625, complement)
p21	Along with CEBP, it can inhibit cdk2 action	11369759	6 NC_000006.11 (36644237..36655116)
Smad3	GR inhibits the transcriptional activation function of Smad3	10518526	15 NC_000015.9 (67358036..67487533)
Smad4	GR inhibits the transcriptional activation function of Smad4 (only in vitro)	10518526	18 NC_000018.9 (48556583..48611412
RanBP9	Overexpression of RanBP9 leads to enhanced GR activity	12361945	6 NC_000006.11 (13621730..13711796, complement)
SET	Acts as ligand-activated GR-responsive transcriptional repressor	18096310	9 NC_000009.11 (131445934..131458675)
NFATc	GR, through protein-protein interaction, interferes with NFATc ability to bind to specific DNA regions	10623828	18 NC_000018.9 (77155772..77289323)
BAFs (BAF57, BAF60a, BAF250a, BAF250b)	Human analogs of the SWI/SNF complex. These complexes partake in glucocorticoid stimulated transcription by interacting with GR.	26278180	17 NC_000017.10 (38781214..38805658, complement) 12 NC_000012.11 (50478760..50494494) 1 NC_000001.10 (27022522..27108601) 6 NC_000006.11 (157098980..157531913)
p53	GR has the ability to inhibit p53-depended functions	22773829 11080152	17 NC_000017.10 (7571720..7590868, complement)
PPP5	Suppression of PP5 results to nuclear accumulation of GR	11389770	19 NC_000019.9 (46850251..46896238)
STAT3	Acts as transcriptional coactivator of the glucocorticoid receptor	9388192	17 NC_000017.10 (40465342..40540586, complement)
STAT5	GR can act as transcriptional coactivator for	8878484	17 NC_000017.10

(STAT5A, STAT5B)	Stat5 and enhance Stat5-dependent transcription		(40439565..40463961) 17 NC_000017.10 (40351195..40428478, complement)
STAT6	Physically and functionally interacts with GR in T-lymphocytes	11150515	12 NC_000012.11 (57489187..57505196, complement)
Thioredoxin(Trx)	Thioredoxin negatively modulates GR function	8958209	9 NC_000009.11 (113006092..113018920, complement)
Mitochondrial Thioredoxin (Trx2)	Mitochondrial thioredoxin has a regulatory role in GR and NFκB signaling pathways. Specifically, Trx2 stimulates the TNFα-induced NFκB activation and DEX-induced GR activation of reporter genes	19570036	22 NC_000022.10 (36863083..36878072, complement)
Thioredoxin reductase 1 (TrxR1)	Overexpression of TrxR1 increases GR activity in specific cells	17382897	12 NC_000012.11 (104609537..104744085)
TRIM28	TRIM28 enhances GR-regulated expression	9742105	19 NC_000019.9 (59055824..59062087)
NCOR1	Represses the GR gene trough a GR-NCOR1-HDAC3 repression complex	23428870	17 NC_000017.10 (15933408..16118874, complement)
HDAC3	Represses the GR gene through a GR-NCOR1-HDAC3 repression complex	23428870	5 NC_000005.9 (141000443..141016423, complement)
NR2F2	NR2F2 represses the GR-stimulated transcriptional activity by tethering corepressors such as NCOR2 and NCOR1. GR stimulates NR2F2 transactivating factors.	15265774	15 NC_000015.9 (96869157..96883492)
NCOR2	Partakes in NR2F2-dependent GR suppression	15265774	12 NC_000012.11 (124808957..125052079, complement)
NFKBIA	GR activates its specific gene to repress NFκB expression	11694573	14 NC_000014.8 (35870716..35873960, complement)
EGFR	GR modulates EGFR function	31052457	7 NC_000007.13 (55086678..55279262)
HMGB1	GR modulates HMGB1 expression	21737101	13 NC_000013.10 (31032877..31191942, complement)
RPS6KA5(MSK1)	Liganded GR interacts with activated RPS6KA5 resulting in redistribution of a part of the nuclear RPSK6KA5 pool to the cytoplasm	20456998	14 NC_000014.8 (91335086..91526993, complement)
Casein kinase 2 (CSNK2A1,	It phosphorylates the Glucocorticoid Receptor	23953592	20 NC_000020.10 (463338..524482,

CSNK2A2, CSNK2B)			complement) 16 NC_000016.9 (58191811..58231782, complement) 6 NC_000006.11 (31632995..31637844)
NLRP3	GR binds on its specific gene and regulates its expression	21940629	1 NC_000001.10 (247579247..247612410)
Mcl-1	GR directly binds on its gene and regulates Mcl-1 expression	20156337	1 NC_000001.10 (150547027..150552214, complement)
NOXA	GR directly binds on its gene and regulates NOXA expression	20156337	18 NC_000018.9 (57567153..57571538)
KLF13	GR binds on the KLF13 promoter to trigger its expression	25336632	15 NC_000015.9 (31619083..31670102)
BIM	Shows an intronic binding site for GR, that is activated in case of dexamethasone sensitivity	25336632	2 NC_000002.11 (111878491..111926022)
FOXO3	GR induces its transcription	22848740	6 NC_000006.11 (108881026..109005972)
BAK	BAK co-precipitates with GR upon dexamethasone treatment	27888447	6 NC_000006.11 (33540323..33548072, complement)
Bcl-xL	Bcl-xL co-precipitates with GR upon dexamethasone treatment	27888447	20 NC_000020.10 (30252261..30311752, complement)
PI3K (p85 subunit /PIK3R1-6)	Physically interacts with GR, they then regulate the tlr2 signaling cascade	19874421	5 NC_000005.9 (67511584..67597649) 19 NC_000019.9 (18263988..18281343 1 NC_000001.10 (46505812..46642167, complement) 3 NC_000003.11 (130397778..130465696, complement) 17 NC_000017.10 (8782233..8869029, complement) 17 NC_000017.10 (8706055..8770994,

			complement)
Annexin1	GR induces its specific gene (ANXA1)	16236742	9 NC_000009.11 (75766721..75785309)
TSLP	GR negatively regulates TSLP's gene expression	23222642	5 NC_000005.9 (110405778..110413722)
ST13	ST13 promotes the functional maturation of GR	27169854	22 NC_000022.10 (41220539..41253012, complement)
PPID	PPID anchors on the GR complex	27169854	4 NC_000004.11 (159630279..159644552, complement)
IRF8	Its' gene is regulated by GR	17185395	16 NC_000016.9 (85932774..85956212)
LAD1	Its' gene is regulated by GR	17185395	1 NC_000001.10 (201349966..201368669, complement)
IGFBP-1	Its' gene is regulated by GR	17185395	7 NC_000007.13 (45927959..45933267)
PKAc (PRKACA PRKACB PRKACG)	GR cross-couples with the catalytic subunit of PKA	27169854	19 NC_000019.9 (14202500..14228559, complement) 1 NC_000001.10 (84543745..84704181) 9 NC_000009.11 (71627426..71635600, complement)
TRIP6	TRIP6 creates a complex with GR, which partakes in the receptor's transrepression ability	27169854	7 NC_000007.13 (100464950..100471076)
14-3-3 (14-3-3σ 14-3-3η 14-3-3ζ/δ)	Takes part in a complex which features GR and Raf-1	27169854	1 NC_000001.10 (27189633..27190947) 22 NC_000022.10 (32340479..32353590) 8 NC_000008.10 (101930804..101965717, complement)
Raf-1	Takes part in a complex which features GR and 14-3-3	27169854	3 NC_000003.11 (12625100..12705700, complement)
PPARγ	Interacts with GR	27169854	3 NC_000003.11 (12329349..12475855)
PPARα	Interacts with GR, and they both act as immunosuppressors	27169854	22 NC_000022.10 (46546458..46639653)

LXR(α,β)	LXR has both synergistic and opposing effects on GR	27169854	11 NC_000011.9 (47269851..47290584) 19 NC_000019.9 (50879680..50886285)
RARα & RXRα	They both bind on GR and enhance its transcriptional activity	27169854	17 NC_000017.10 (38465423..38513895) 9 NC_000009.11 (137218316..137332431)
Progesterone Receptor	They possibly interact to repress IL-1β-driven COX-2 activation	27169854	11 NC_000011.9 (100900355..101000544, complement)
Estrogen Receptor alpha	Its interaction with GR can have cooperative or antagonistic action on E2-regulated genes	27169854	6 NC_000006.11 (152011631..152424409)
Nur77	Through protein-protein interaction GR antagonizes Nur77-depenent transcription on the Nur77 response element of the pomc gene	27169854	12 NC_000012.11 (52416616..52453291)
SOCS1	GR ad SOCS1 create an intracellular complex and GCs increase the nuclear levels of SOCS1	18524780	16 NC_000016.9 (11348274..11350039, complement)
Tbx21/T-bet	GR interacts with Tbx21 and inhibits Tbx21's action	27169854	17 NC_000017.10 (45810610..45823485)
FOXA3	FOXA3 mediates GR function in adipose tissue	26957608	19 NC_000019.9 (46367518..46377055)
PER2	GR regulates its function	19805059	2 NC_000002.11 (239152679..239198678, complement)
RSUME	Possibly interacts with GR and takes part in the receptor's sumoylation	27169854	1 NC_000001.10 (95699711..95712781)
SUMO4	SUMO4-induced GR sumoylation enhances GR DNA binding activity	27169854	6 NC_000006.11 (149721284..149722182)
Ubch7	It interacts with GR and its effects on the receptor depend on the cell culture studied	27169854	22 NC_000022.10 (21903736..21978323)
E6-AP	E6-AP regulates GR transactivation	27169854	15 NC_000015.9 (25582394..25684190, complement)

Each gene region was then located in the dataset based on their genetic locus and all relative SNPs were extracted. A sliding window algorithm was then used to retrieve the regions of interest with all the identified SNPs which have a reference SNP ID number and are present in the dbSNP database. All the extracted SNPs were later stored in a structured database and included relevant information extracted from the primary dataset, such as gene name, genetic position,and change and frequency of occurrence based on the sample.

The extracted SNPs were updated to the current GRCh38.p13 assembly and later annotated with relative information from the dbSNP database, clinVar database (255) and LitVar database (214) **(Figure 20)**. By using a set of rules based on each database protocols, several types of information were extracted and included in the resulting database. Particularly, the clinVAR database was used to find possible associations with human health, the dbSNP database to find if the SNPs type, common changes and the allele frequency in different populations and the LitVar database to find the most co-occurred literature entries regarding diseases, chemicals and variants. Based on the results and the available information received through the aforementioned annotation process, an ontology analysis was performed and the SNPs were evaluated based on their availability in the ClinVar database and the corresponding information present in LitVar. This analysis was conducted in order to display a general picture of the most studied mechanisms that the GR interactome may play a part. Finally, summarizing all the information collected for each SNPs, a comparison with a dataset of 1465 Korean individuals was conducted in an attempt to identify characteristics specific to the Japanese population that are associated with the GR interactome. The comparison of these two populations was chosen due to their genetic similarities (256).

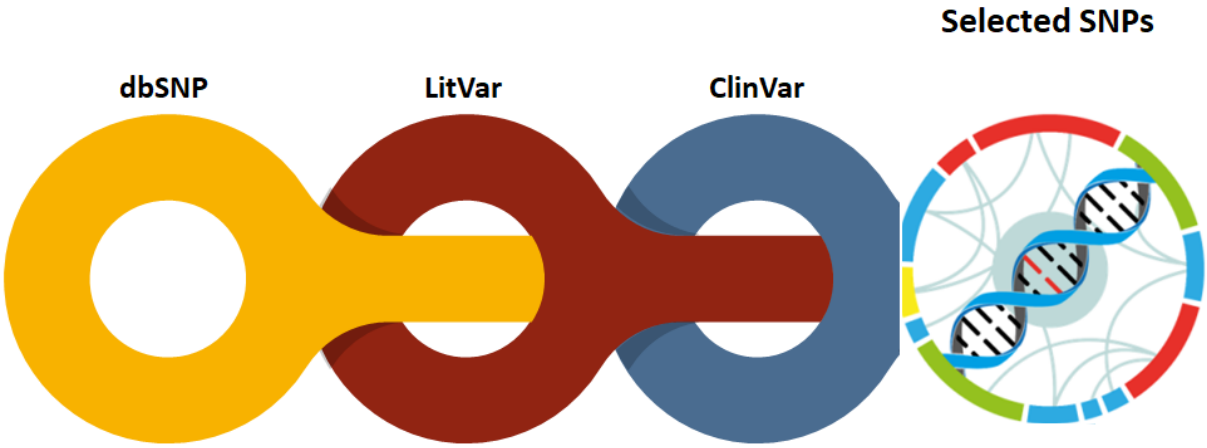

Figure 20 Information extraction from the dbSNP, LitVar, and ClinVar databases

Results

The genes checked amounted to 31600 SNPs with a known rs ID that were present in the dbSNP database. Out of the above, 411 SNPs were present in the ClinVar database and were chosen as possible SNPs of interest, while an ontology analysis based on the corresponding LitVar entries was conducted on these SNPs in an effort to paint a general picture of the GR interactome literature. Most studies regarding the GR interactome seem to focus on neuropsychiatric and metabolic disorders, including drug metabolism, plus various neoplasms **(Figure 21)**. An interesting inclusion is also Zellweger Syndrome, a characteristic peroxisomal biogenesis disorder that emerges due to PEX genes mutations. These mutations lead to various metabolic abnormalities, with resulting symptoms being liver dysfunction,

neurological abnormalities such as developmental delay, adrenocortical dysfunction, in addition to vision and hearing impairment. GR is an important mediator of proper adrenocortical function and cortisone supplementation is often used as a therapeutic approach in severe cases of Zellweger syndrome (257). Metabolism and psychiatric disorders are apparent fields of study when it comes to glucocorticoid receptor function, but the extremely high prevalence of cancer research is a -somewhat- unexpected result. As it has been stated, the glucocorticoid receptor's role in cancer seems to be context-dependent (258) and thus the use of this receptor as a potential drug target or biomarker is a complex affair. Nonetheless, since cancer research has rapidly increased over the past decades (259), scientists have started focusing on more complex regulators of cancer such as GR, which may explain the high number of cancer studies regarding the GR interactome.

Figure 21 Ontology analysis of the 411 SNPs present in the ClinVar database based on LitVar-received information

Out of the 411 polymorphisms, 46 SNPs showcased a known association with human health according to the clinVAR database **(Table 10)**. The vast majority of these SNPs were associated with drug response and metabolic disorders, something not surprising since GR has an essential role in metabolism and cytochrome P450 function (260, 261). An interesting observation is that many drugs whose metabolism is influenced by the GR interactome are anti-depressants, which is in accordance with GR's role in neuropsychiatric disorders pathology and particularly depression (92). Lastly, an association with diseases such as chronic obstructive pulmonary disease and inflammatory bowel disease was present in the SNPs, which are characteristic inflammatory diseases, where glucocorticoids are used as potent anti-inflammatory medication (262).

Table 10 A list of GR interactome SNPs that have an effect on human health according to the clinvar database.

Gene	SNP	Nucleotide change	Ref allele freq	Alt allele freq	Association
HSPA1L	rs2227956	G>A	0,0858	0,9142	Chronic Obstructive Pulmonary Disease
HSPA1L	rs2227955	T>G	0,9799	0,0201	Inflammatory Bowel Disease
HSPA1L	rs34620296	C>T	0,9983	0,0017	Inflammatory Bowel Disease
HSPA1L	rs368138379	C>T	0,9999	0,0001	Inflammatory Bowel Disease
HSPA1B	rs6457452	C>T	0,9378	0,0622	Chronic Obstructive Pulmonary Disease
HSPA1A	rs1043618	G>C	0,8401	0,1599	Chronic Obstructive Pulmonary Disease
CYP3A5	rs4646450	G>A	0,7399	0,2601	Appendicular Lean Mass Relative to Body Height
TP53	rs201753350	C>T	0,9938	0,0062	Li-Fraumeni syndrome 1
FKBP5	rs4713916	A>G	0,1999	0,8001	Influences Efficacy of Antidepressants (Citalopram, Fluoxetine, Mirtazapine, Paroxetine, SSRIs, Venlafaxine)
CYP2C9	rs1057910	A>C	0,9758	0,0242	Influences Warfarin Metabolism
CYP2C9	rs7089580	A>T	0,9900	0,0100	Influences Warfarin Response
CYP2C9	rs4917639	A>C	0,8526	0,1474	Influences Warfarin Response
CYP2C19	rs4244285	G>A	0,7056	0,2944	Influences Clopidogrel Response (Efficacy, Toxicity/ADR); Influences Clomipramine Response (Efficacy); Influences Amitriptyline Response (Efficacy); Influences Citalopram Response (Efficacy); Poor Metabolism of Mephenytoin; Poor Metabolism of Proguanil; Poor Metabolism of Clopidogrel
CYP3A5	rs776746	T>C	0,2444	0,7556	Influences Tacrolimus response based on Recipient Genotype (Dosage, Metabolism/PK);Influences Sirolimus Response (Metabolism/PK);Influences Cyclosporine Response (Dosage, Metabolism/PK);Influences Tacrolimus Response based on Donor Genotype (Dosage, Metabolism/PK); Influences Tacrolimus Response

					(Efficacy);Influences Sirolimus Response (Dosage)
CYP2C19	rs72552267	G>A	0,9997	0,0003	CYP2C19:No Function
ABCB1	rs1045642	A>G	0,4119	0,5881	Influences Fentanyl Response (Efficacy); Influences Methadone Response (Dosage, Efficacy); Influences Morphine Response (Dosage, Efficacy); Influences Opiods Response (Dosage,Efficacy); Influences Oxycodone Response (Dosage, Efficacy);Influences Tramadol Response (Dosage, Efficacy); Influences Tramadol Response;Influences Nevirapine Response (Toxicity/ADR); Influences Digoxin Response (Toxicity/ADR); Influences Ondansetron Response (Efficacy);Influences Methotrexate Response (Toxicity/ADR)
ABCB1	rs3842	T>C	0,7203	0,2797	Influences Tramadol Response
ABCB1	rs1922242	A>T	0,6649	0,3351	Influences Tramadol Response
ABCB1	rs2235046	T>C	0,6052	0,3948	Influences Tramadol Response
ABCB1	rs2235013	C>T	0,6167	0,3833	Influences Tramadol Response
ABCB1	rs2235035	G>A	0,6813	0,3187	Influences Tramadol Response
ABCB1	rs2235033	A>G	0,6294	0,3706	Influences Tramadol Response
ABCB1	rs139611979	C>T	0,9992	0,0008	Influences Tramadol Response
ABCB1	rs10276036	C>T	0,6184	0,3816	Influences Tramadol Response
ABCB1	rs1922240	T>C	0,6840	0,3160	Influences Tramadol Response
ABCB1	rs28381877	A>G	0,9999	0,0001	Influences Tramadol Response
ABCB1	rs868755	T>G	0,4118	0,5882	Influences Tramadol Response
ABCB1	rs13237132	C>G	0,6832	0,3168	Influences Tramadol Response
ABCB1	rs1202170	C>T	0,3856	0,6144	Influences Tramadol Response
ABCB1	rs1202168	G>A	0,3846	0,6154	Influences Tramadol Response
ABCB1	rs1016793	G>A	0,5916	0,4084	Influences Tramadol Response
ABCB1	rs2235018	T>C	0,7931	0,2069	Influences Tramadol Response
ABCB1	rs28381827	C>T	0,8748	0,1252	Influences Tramadol Response
ABCB1	rs1211152	A>C	0	1	Influences Tramadol Response
ABCB1	rs373236080	C>T	0,9999	0,0001	Influences Tramadol Response
ABCB1	rs2235074	G>A	0,9291	0,0709	Influences Tramadol Response
ABCB1	rs2214102	T>C	0	1	Influences Tramadol Response
ABCB1	rs3213619	A>G	0,9289	0,0711	Influences Tramadol Response
VDR	rs2228570	A>G	0,3674	0,6326	Influences Response to Peginterferon Alfa-2b and

					Ribavirin (Efficacy)
FKBP5	rs1360780	T>C	0,2246	0,7754	Major Depressive Disorder; Increased Recurrence of Depressive Episodes; Susceptibility to Major Depressive Disorder; Accelerated Response to Antidepressant Drug Treatment
SUMO4	rs237025	G>A	0,3028	0,6972	Type 1 Diabetes Mellitus
PPARG	rs28936407	G>A	0,9999	0,0001	Somatic Colon Cancer
PPARG	rs1801282	C>G	0,9695	0,0305	Type 2 Diabetes melitus
TAT	rs118203914	G>A	0,9999	0,0001	Tyrosinemia Type 2
PPARA	rs1800206	C>G	0,9999	0,0001	Susceptibility to Hyperapobetalipoproteinemia
SMAD4	rs12456284	A>G	0,5757	0,4243	Confers sensitivity to lung cancer

The 46 SNPs above were then checked on the LitVar database in an attempt to gain more information regarding their role in GR signaling and homeostasis **(Table 11)**. Out of those ClinVar variations, four ABCB1 variations did not display a corresponding LitVar entry. Those variations are rs373236080, rs28381827, rs28381877 and rs139611979 with such a discrepancy possibly emerging because ClinVar also includes information beyond literature-described associations, like lab data (263). The SNPs which displayed both an entry in LitVar and an entry with a possible pathological association in ClinVar, were termed SNPs of interest. The results are in line with the ClinVar-obtained information. Some novel associations with various neoplasms seem to emerge, though those associations are mainly limited to ABCB1 SNPs and since that gene codes for the P-glycoprotein, which has an extensively studied role in cancer multidrug resistance (264), such results are expected.

Table 11 A list featuring SNPs of interest, number of publications attributed to each SNP, which gene they are located in, plus diseases, chemicals, and variants that most commonly co-occur with mentioned SNPs in the same sentence

S/N	SNPs	Pub.	Located in Gene	Variants Co.	Diseases Co.	Chemicals Co.
1	rs4713916	49	FKBP5	rs1360780 rs3800373 rs9470080 rs9296158 rs41423247 rs4713902 rs7997012 rs9394309 rs6265	-Depressive Disorder -Major Depressive Disorder -Wounds and Injuries -Abusive Dwarfism Syndrome	-hydrocortisone -Citalopram -Serotonin -Ethanol -C081489 -Steroids

					-Anxiety Disorders	
2	rs1360780	271	FKBP5	rs3800373 rs9296158 rs9470080 rs4713916 rs6265	-Depressive Disorder -Post Traumatic Stress Disorder -Wounds and Injuries -Major Depressive Disorder -Mental Disorders	-Hydrocortisone -Ethanol -Dexamethasone -Tacrolimus
3	rs1045642	1984	ABCB1	rs1128503 rs2032582 c.2677G>T,A rs776746 rs2231142 rs4244285 rs1801133 rs717620 rs4149056	-Epilepsy -Breast Neoplasms, -Neoplasms -Drug-Related Side Effects and Adverse Reactions -Colorectal Neoplasms	-Clopidogrel -Tacrolimus -Cyclosporine -Digoxin -Methotrexate -Peptide T amide
4	rs3842	26	ABCB1	rs1045642 rs3745274 rs776746 rs1128503 rs10264272	-Hypertension -Diabetes Mellitus -Lung Neoplasms -Dyslipidemias -HIV Infections	-Efavirenz -1-chloro-2-hydroxy-3-butene -Carbon -1,7,9,11-tetrahydroxy-3-methyl-8,13-dioxo-5,6,8,13-tetrahydrobenzo(a)tetracene-2-carboxylic acid
5	rs1922242	8	ABCB1	rs1045642 rs1202184 rs10808072 rs3213619 rs1128503 rs2032582 rs1202168	-Renal Cell Carcinoma -Depressive Disorder -Anxiety Disorders -Seizures	-12-(4'-azido-2'-nitrophenoxy)dodecanoyl-coenzyme A -Oxygen -Thulium
6	rs2235046	9	ABCB1	rs1128503 rs10276036 rs1202169 rs4148738 rs1045642	-Renal Insufficiency -Zellweger Syndrome -Lung Neoplasms -N syndrome -Bilateral Multicystic Renal Dysplasia	-C554682 -Apixaban -C065179 -Nitrogen -C503223 -Interleukin-2 Receptor beta Subunit -Carbon

7	rs2235013	8	ABCB1	rs1045642 rs1128503 rs2235033 rs2032582 rs1202179 rs1695 rs10276036 rs2235046 rs9282564	-Follicular Thyroid Cancer -Lung Neoplasms -Proteinuria -Ataxia Telangiectasia -Zellweger Syndrome	-Cyclosporine -Pentalysine -Alanyl-alanyl-alanyl-alanine -Methionylmethionine -Seryl-seryl-seryl-arginine -Leucylleucine -Peptide T amide -2'-deoxy-5-fluoro-3'-thiacytidine -Carbon -Interleukin-2 Receptor beta Subunit
8	rs2235035	6	ABCB1	rs1202169 rs2032582 rs1045642 rs1138272 rs4520 rs1292798 rs4891 rs1027649 rs2235046	-N syndrome -Ataxia Telangiectasia	-Alanyl-alanyl-alanyl-alanine -Triglycerides -Angoletin -Seryl-seryl-seryl-arginine -Peptide T amide
9	rs2235033	6	ABCB1	rs1128503 rs1045642 rs2235013 rs2235046 rs4148738 rs4680 rs2273697 rs10276036 rs4437575	-Zellweger Syndrome	-Sulfur -Carbon -Interleukin-2 Receptor beta Subunit -Daunorubicin -Cyclosporine -Daunorubicinol
10	rs10276036	15	ABCB1	rs1128503 rs2235046 rs1202169 rs1202167 rs1202168 rs4148738	-Neutropenia -Diabetes Mellitus -Hypertension -Breast Neoplasms -Neoplasms	-C554682 -Doxorubicin -Glycyl-glycyl-sarcosine -Warfarin -Serotonin -Apixaban -Irinotecan -Superoxides -Adenosine triphosphate -Guanosine
11	rs1922240	2	ABCB1	rs6591256 rs1338062 rs754814 rs7793196 rs7223183 rs11869640 rs4148732 rs7299040	-Pain -N-Syndrome	-Morphine

				rs5993875		
12	rs868755	9	ABCB1	rs1858923 rs1202168 rs1045642 rs10280623 rs4148738 rs2032582 rs7779562 rs10808072 rs2235048	-Drug-Related Side Effects and Adverse Reactions - Zellweger Syndrome - Colorectal Neoplasms	-Carbon -Interleukin-2 Receptor beta Subunit
13	rs13237132	2	ABCB1	rs2235023 rs4148732 rs12334183 rs10264990 rs238416 rs12129768 rs11188148 rs1678607 rs50872	- Ovarian Neoplasms - Bradycardia -Von Hippel-Lindau Disease	-
14	rs1202170	1	ABCB1	rs1799971 rs2235033 rs3024971 rs3786047 rs2235013 rs167769 rs1045642 rs1045280 rs2036657	-	-
15	rs1202168	11	ABCB1	rs1045642 rs1128503 rs2032582 rs868755 rs1202169 rs2235046 rs10276036 rs4148738 rs1202167	- Colorectal Neoplasms - Zellweger Syndrome -Neonatal Hyperbilirubinemia	-C554682 -Thulium -C065179 -Apixaban -12-(4'-azido-2'-nitrophenoxy)dodecanoyl-coenzyme A -Interleukin-2 Receptor beta Subunit -Carbon -Oxygen -C503223
16	rs1016793	2	ABCB1	rs6961665 rs41277128 rs62578960 rs10985911 rs1202168 rs116855710 rs114717568 rs11673270	-Zellweger Syndrome	-Carbon -Interleukin-2 Receptor beta Subunit

				rs11083571		
17	rs2235018	1	ABCB1	rs34800935 rs7793933 rs2188526 rs1045642 rs6949448 rs1922244 rs2235048 rs4148738 rs7787082 rs12720464	-	-Carbon
18	rs1211152	3	ABCB1	rs1045642 rs10264990 rs1202184 rs17327624 rs6946119	-	-
19	rs2235074	2	ABCB1	rs35979566 rs2279342 rs1202169 rs6591722 rs1010570 rs4244285 rs4148329 rs1042838 rs7801671	-Myelodysplastic Syndromes	-Adenosine Triphosphate -Nitrogen
20	rs2214102	12	ABCB1	rs1045642 rs1128503 rs2229109 rs2032582 rs3213619 rs9282564	-Breast Neoplasms -Ataxia Telangiectasia	-Decaglycine -Peptide T amide -His-His-His-His-His-His -Seryl-seryl-seryl-arginine -Leucylleucine -Triamcinolone -Progestins -Diprotin A -Asparagyl-aspargyl-tryptophyl-asparagyl-asparagine -1,7,9,11-tetrahydroxy-3-methyl-8,13-dioxo-5,6,8,13-tetrahydrobenzo(a)tetracene-2-carboxylic acid
21	rs3213619	75	ABCB1	rs1045642 rs1128503 rs2032582 c.2677G>T,A rs776746	-Colorectal Neoplasms -Hypertension -Drug-Related Side Effects and Adverse Reactions	-Tacrolimus -Paclitaxel -Docetaxel -C097613 -Vasoactive intestinal constrictor -Taxane -Methotrexate-alpha-phenylalanine

					-Dyslipidemias -Diabetes Mellitus	-Carbon -Decaglycine -Cyclosporine
22	rs4986893	276	CYP2C19	rs4244285 rs12248560 rs1057910 rs1799853 rs1045642	-Drug-Related Side Effects and Adverse Reactions -Hypertension -Breast Neoplasms -Thrombosis -Stroke	-Clopidogrel -Warfarin -Simvastatin -Carbon -Tryptophyl-arginyl-tryptophyl-tryptophyl-tryptophyl-tryptophanamide
23	rs4244285	475	CYP2C19	rs4986893 rs12248560 rs1057910 rs1799853 rs1045642	-Drug-Related Side Effects and Adverse Reactions -Blood Platelet Disorders -Breast Neoplasms -Hemorrhage -Thrombosis	Clopidogrel Warfarin Simvastatin Aspirin Carbon
24	rs72552267	31	CYP2C19	rs41291556 rs28399504 rs4986893 rs56337013 rs4244285	-N syndrome -Abnormal Reflex -Norrie Disease -Drug-Related Side Effects and Adverse Reactions -Acute Coronary Syndrome	-Clopidogrel -Warfarin -Simvastatin -Metformin -Tacrolimus
25	rs1057910	512	CYP2C9	rs1799853 rs9923231 rs4244285 rs2108622 rs4986893	-Drug-Related Side Effects and Adverse Reactions -Hemorrhage -Diabetes Mellitus -Neoplasms -Hypertension	-Warfarin -Clopidogrel -Simvastatin -Ciproxifan -Phenytoin
26	rs7089580	12	CYP2C9	rs61162043 rs9923231 rs1799853 rs7900194 rs28371686 rs28371685	-Atrial Fibrillation -Norrie Disease	-Warfarin -Amiodarone -Carcinoma-associated --Antigen 17-1A -Synthetic SNP-1 protein

				rs1057910		
27	rs4917639	18	CYP2C9	rs1057910 rs9923231 rs1799853 rs7294 rs10871454	-Stroke -Zellweger Syndrome -Intracranial Hemorrhages	-Warfarin -C065179 -Carvedilol -S-imvastatin -Acenocoumarol -Metoprolol -Pravastatin -Vitamin K1 oxide -Clopidogrel -Sodium
28	rs4646450	12	CYP3A5	rs15524 rs776746 rs4244285 rs1128503 rs1045642	-Cardiotoxicity -Urinary Bladder Neoplasms -Graft vs Host Disease	-12-(4'-azido-2'-nitrophenoxy)dodecanoyl-coenzyme A -Tacrolimus -Aldrin -Lopinavir -Sch 601324 -Alachlor -Cyanazine -Calcium -Poly(acrylamide-co-crotonic acid) -Dehydroepiandrosterone Sulfate
29	rs776746	524	CYP3A5	rs1045642 rs2740574 rs1128503 rs2032582 rs35599367	-Drug-Related Side Effects and Adverse Reactions -Hypertension -Neoplasms -Diabetes Mellitus -Non-Small-Cell Lung Carcinoma	-Tacrolimus -Simvastatin -Cyclosporine -Clopidogrel -Warfarin -Sunitinib
30	rs1043618	59	HSPA1A	rs2227956 rs1061581 rs1008438 rs2075800 rs2763979	-Major Depressive Disorder -Alzheimer Disease -Depressive Disorder -Glaucoma -Neoplasms	-2-carboxyarabinitol 1-phosphate Gastrofenzin Poly-aluminum-chloride-sulfate Nitroglycerin Glycyl-threonine Oxytocin, Glu(4)- Hydrogen Nitrogen Tyrosyl-lysine -Human LRRN2 protein
31	rs6457452	11	HSPA1B	rs2763979 rs1061581 rs17200983 rs13118	-Alopecia Areata -Schizophrenia -Paranoid Schizophrenia	-15-hydroxy-5,8,11,13-eicosatetraenoic acid -3,4,5-trichloroguaiacol Cholesterol

				rs150142878 rs9267546 rs11538264 rs9267547 rs4576240	-Anemia -Malaria	-Methacholine Chloride -Triglycerides -Prostaglandins -Uric Acid -Aspirin -Carbon
32	rs2227956	63	HSPA1L	rs1061581 rs1043618 rs2075800 rs2763979 rs662	-Stomach Neoplasms -Male Infertility -Neoplasms -Ataxia Telangiectasia -Diabetic Foot	-Peptide T amide -Methionylmethionine -Glycyl-threonine -Methionine -1,10-phenanthroline-5,6-dione
33	rs2227955	2	HSPA1L	rs2227956 rs2075800 rs35326839 rs10117 rs116768554 rs14355 rs566393477 rs1042881 rs34620296	-Ataxia Telangiectasia	-Alanine -Glycine -1,10-phenanthroline-5,6--dione -Peptide T amide -Glycyl-threonine -Methionine -Arginyl-glutamine
34	rs34620296	2	HSPA1L	rs139193421 p.K73S rs2075799 rs368138379 rs199780750 rs139868987 rs2227956 c.515_517del c.218A>G	-Multiple Hamartoma Syndrome -Crohn Disease -Proctitis -Colitis -Gastritis	-1,10-phenanthroline-5,6-dione -IS 23 -Peptide T amide
35	rs368138379	1	HSPA1L	rs199780750 rs2075799 rs116768554 rs1460318497 rs566393477 rs35326839 rs2075800 p.172del rs35347921 rs9469057	-Multiple Hamartoma Syndrome -Crohn Disease -Proctitis -Colitis -Gastritis	-1,10-phenanthroline-5,6-dione -Peptide T amide
36	rs2279744	373	MDM2	rs1042522 rs117039649 rs1801270 rs25487 rs9344	-Neoplasms -Lung Neoplasms -Breast Neoplasms -Stomach	Arginylarginine Estrogens Synthetic SNP-1 protein Gastrofenzin Nitrogen

					Neoplasms -Endometrial Neoplasms	Cisplatin
37	rs2228570	717	VDR	rs1544410 rs731236 rs7975232 rs11568820 rs7041	-Ovarian Neoplasms -Asthma -Breast Neoplasms -Neoplasms -Multiple Sclerosis	-Vitamin D -25-hydroxyvitamin D3-bromoacetate -Calcium -Poly If -Peptide T amide
38	rs201753350	18	TP53	rs28934576 rs1042522 rs1800371 rs1800370 rs730882025 rs28934578 rs1800372 rs104886003 rs1057519991	-Rhabdomyosarcoma -Emanuel Syndrome -Acute Myeloid Leukemia -Neoplasms -Norrie Disease	-AT 61 -Arginyl-tryptophyl-arginine Nitrogen Histocompatibility Antigen H-2D Leucylleucine -3-bromoacetoxyandrostan-17-one -2-(3,4-dimethoxyphenyl)-5-amino-2-isopropylvaleronitrile -chromozym TH -(arginine)9-cysteinyl-glutaminyl-cysteinyl-arginyl-arginyl-lysyl-asparagine -H 189
39	rs1800206	217	PPARA	rs2016520 rs1801282 rs4253778 rs3856806 rs135539 rs1805192	-Diabetes Mellitus -Obesity -Metabolic Diseases -Type 2 Diabetes Mellitus -Atherosclerosis	-Triglycerides - Omega-3 Fatty Acids -Fatty Acids -Cholesterol -Unsaturated Fatty Acids
40	rs1801282	979	PPARG	rs7903146 rs5219 rs13266634 rs4402960 rs10811661 rs1111875 rs3856806 rs864745 rs7961581	-Diabetes Mellitus -Obesity -Type 2 Diabetes Mellitus -Metabolic Diseases -Neoplasms -Hypertension -Insulin Resistance -Coronary Disease -Colorectal Neoplasms	-Alanyl-alanyl-alanyl-alanine -N-nitroso-prolylalanine -Glucose -Troglitazone -Thiazolidinediones -Cholesterol -Triglycerides -Ethanol -Potassium -Fatty Acids

					-Polycystic Ovary Syndrome	
41	rs28936407	6	PPARG	rs121909242 rs72551362 rs121909243 rs72551364 rs72551363 rs121909244	-Neoplasms -Colorectal Neoplasms -Lipodystrophy -Lipid Metabolism Disorders -Migraine Disorders	-Urea -Hydrogen -Rosiglitazone -Lecithin emulsion safflower oil
42	rs237025	132	SUMO4	rs2476601 rs577001 rs1805010 rs237024 rs1800872 rs2243250	-Diabetes Mellitus -Type 1 Diabetes Mellitus -Type 3 Axenfeld-Rieger syndrome -Type 2 Diabetes Mellitus -Diabetic Nephropathies	-Methionine -Valine-valine-saquinavir Tacrolimus -(Z)-2-amino-5-chlorobenzophenonamidinohydrazone acetate -Nitrogen -Triglycerides
43	rs118203914	1	TAT	p.E411X (1) rs761817519 (1) p.L201R (1) rs758306831 (1) rs775488556 (1)	Type 2C Congenital Disorder of Glycosylation	-

The SNPs of interest are then characterized based on nucleotide change, nucleotide change region, and mutation type. The Japanese dataset is then compared to a dataset of Korean individuals focusing on the frequencies exhibited by the SNPs of interest **(Table 12)**. The vast majority of SNPs of interest are located in intronic regions, which play an important role in gene regulation. The comparison among the two populations pinpointed the rs1043618 as a polymorphism with a considerably different frequency among Japanese and Korean individuals. This polymorphism has been associated with depression in LitVar and COPD in response to environmental factors in ClinVar.

Table 12 A comparison of SNP frequencies among Korean and Japanese individuals

Gene	SNP	Nucleotide change	Type of mutation/ Nucleotide change region	Nucleotide frequency in the Japanese population	Nucleotide frequency in the Korean population

HSPA1L	rs2227956	G>A	Missense variant	G=0,0858	G=0,0765
HSPA1L	rs2227955	T>G	Missense variant	G=0,0201	G=0,0171
HSPA1L	rs34620296	C>T	Missense variant	T=0,0017	T= 0,0048
HSPA1L	rs368138379	C>T	Missense variant	T=0,0001	-
HSPA1B	rs6457452	C>T	5 Prime UTR Variant	T=0,0622	T=0,0875
HSPA1A	rs1043618	G>C	5 Prime UTR Variant	C=0,1599	C=0,2801
CYP3A5	rs4646450	G>A	Intron Variant	A=0,2601	A=0,2304
TP53	rs201753350	C>T	Missense Variant	T=0,0062	T=0,0055
FKBP5	rs4713916	A>G	Intron Variant	A=0,1999	A=0,2096
CYP2C9	rs1057910	A>C	Missense Variant	C=0,0242	C=0,0413
CYP2C9	rs7089580	A>T	Intron Variant	T=0,01	T=0,0082
CYP2C9	rs4917639	A>C	Intron Variant	C=0, 1474	C=0,1345
CYP2C19	rs4244285	G>A	Synonymous Variant	A=0,2944	A=0,2765
CYP3A5	rs776746	T>C	Splice Acceptor Variant	C=0,2444	C=0,2249 (1K)
CYP2C19	rs72552267	G>A	Missense Variant	A=0,0003	-
ABCB1	rs1045642	A>G	Missense Variant	A=0,4119	A=0,3488
ABCB1	rs3842	T>C	3 Prime UTR Variant	C=0,2797	C=0,3061
ABCB1	rs1922242	A>T	Intron Variant	T=0,3351	T=0,3717
ABCB1	rs2235046	T>C	Intron Variant	C=0,3948	C=0,4085
ABCB1	rs2235013	C>T	Intron Variant	T=0,3833	T=0,4065
ABCB1	rs2235035	G>A	Intron Variant	A=0,3187	A=0,3590
ABCB1	rs2235033	A>G	Intron Variant	G=0, 3706	G=0,4065
ABCB1	rs10276036	C>T	Intron	T=0,3816	T=0,4061

			Variant		
ABCB1	rs1922240	T>C	Intron Variant	C=0,3160	C=0,3573
ABCB1	rs868755	T>G	Intron Variant	T=0,4118	T=0,3788
ABCB1	rs13237132	C>G	Intron Variant	G=0,3168	G=0,3563
ABCB1	rs1202170	C>T	Intron Variant	C=0,3856	C=0,4058
ABCB1	rs1202168	G>A	Intron Variant	G=0,3846	G=0,4038
ABCB1	rs1016793	G>A	Intron Variant	A=0,4084	A=0,3860
ABCB1	rs2235018	T>C	Intron Variant	C=0,2069	C=0,2160
ABCB1	rs1211152	A>C	Intron Variant	A=0	A=0,001
ABCB1	rs2235074	G>A	Intron Variant	A=0,0709	A=0,0565
ABCB1	rs2214102	T>C	Synonymous Variant	T=0	T=0,0003
ABCB1	rs3213619	A>G	Intron Variant	G=0,0709	G=0,0561
VDR	rs2228570	A>G	Initiator Codon Variant	A=0,3674	A=0,4041
FKBP5	rs1360780	T>C	Intron Variant	T=0,2246	T=0,2392
SUMO4	rs237025	G>A	Missense Variant	G=0,3028	G=0,2973
PPARG	rs28936407	G>A	Missense Variant	A=0,0001	-
PPARG	rs1801282	C>G	Missense Variant	G=0,0305	G=0,0517
TAT	rs118203914	G>A	Stop Gained	A=0,0001	-
PPARA	rs1800206	C>G	Missense Variant	G=0,0001	G=0,0005(1K)
SMAD4	rs12456284	A>G	3 Prime UTR Variant	0,4243	G=0,4049

Discussion

Polymorphisms on genes characteristic of the GR interactome lead to psychological and inflammatory diseases. These results are in accordance with the stress response system's role in neuropsychiatric disorders (265) and the important role of the glucocorticoid receptor in inflammation (266). Additionally, the importance of

genetic alterations in intronic regions was also highlighted since the majority of variants associated with pathologic conditions were located in introns. Moreover, the genetic similarity between the Korean and Japanese populations is in accordance with prior research, which states that a human migration wave from the Korean Peninsula to Kyushu, the most southerly of Japan's large islands, around 3000 years ago, played a major role in the genetic composition of the current Japanese population (256). Nonetheless, an interesting discrepancy was present between these two populations, which extended to discrepancies with the frequencies present on the TOPMED (267) and 1000 Genomes Project (268). Japanese individuals showcased an rs1043618 frequency of 0.1599 while Koreans had a frequency of 0.2801, with the TOPMED and 1000 Genomes Project frequencies being 0.478474 and 0.4812, respectively. According to the ClinVar database, this HSPA1A polymorphism has been associated with susceptibility to chronic obstructive pulmonary disease in response to environmental stressors in a Mexican population(269). This observation is really intriguing, since COPD displays a higher incidence rate in Korean than in Japanese individuals, with smoking habits being pretty similar among these populations (270, 271). HSP1A codes for Hsp70, an essential regulator of GR signaling. Hsp70 plasma levels have already been associated with COPD (272). Additionally, increased levels of Hsp70 may weaken a COPD patient's response to glucocorticoids (273). This may lead to the speculation that the rs1043618 could be partially responsible for such a phenomenon. Nevertheless, it is important to state that specific SNPs may be associated with a disease in one population but show no association in another one (274).

Conclusions

Gene expression, i.e., the process in which information encoded in a gene is converted into a functional gene product, is considered one of the underlying factors for the emergence of phenotypic variation. Understanding the intricacies of phenotypic variation can elucidate why specific phenotypes display a higher risk of disease or why disease phenotypes themselves seem to differ among individuals wildly. One of the most significant groups of gene expression regulators are transcription factors, which are proteins that bind to regulatory regions in the DNA and promote or inhibit gene transcription. Genetic or structural alterations on transcription factors have played an important role in both the evolution of complex gene regulatory networks and the emergence of pathologic phenotypes. One of the largest families of transcription factors are nuclear receptors, whose function is ligand-dependent and govern biological mechanisms of great importance, including homeostasis maintenance, i.e., the preservation of an inner equilibrium. Since ligand-binding is essential in the evolution of nuclear receptors and their characteristics, researching the evolution of the structural regions that govern the ligand-receptor interaction may help provide novel information regarding gene regulatory networks and, subsequently, phenotypic variation.

Therefore, a hybrid phylogenetic analysis on nuclear receptors ligand-binding domain was conducted. The results highlighted that mutations on interaction sites

are somewhat frequent, but mutations on sites that maintain LBD structure are rarer. It appears that structure is essential for proper function, and mutations on interaction sites that do not influence structure are at the forefront of NR evolution. Four distinct monophyletic branches were uncovered while two major canonical forms were present, one steroid-hormone-like and one thyroid-hormone-like.

Another finding of this analysis was that the estrogen receptor alpha receptor seemed to split into two distinct subcategories based on its structure. These subcategories sequence differences were almost non-existent. It appears that the breast-cancer-associated Y537S mutation may be the main reason for this discrepancy. This mutation seems to lead to a conformational change that gives ERa the ability to bind both ERa and ERb ligands indiscriminately. The hybrid phylogenetic study also showed that the NR-box, a motif mainly found on NR cofactors, is also found in nuclear receptors. This motif mediates interaction with nuclear receptors, and its presence on NRs themselves indicates that interaction among receptors may be more common than expected. Out of all nuclear receptors, the glucocorticoid receptor is most associated with homeostasis maintenance. The ligand-binding domain of the glucocorticoid receptor is characteristic of steroid hormone receptors, showcasing distinct signaling motifs and structures. Research on this receptor's literature could provide information that may also apply to several steroid hormone receptors or even nuclear receptors.

A thorough search on GR and GR signaling modulators literature mostly displayed the known importance of NRs and specifically GR signaling in numerous biological processes like immune response, metabolism, development, and proper brain function. The literature was focused on some intriguing subjects, such as the POLR1C gene, several intronic regions, and cancer. Intronic regions are known to produce non-coding RNAs, which have a significant role in gene expression. POLR1C codes for a subunit of RNA polymerase I and RNA polymerase III. Both enzymes participate in rRNA synthesis, and stress influences this process. The research on GR's role in pathways regulating rRNA synthesis is not that extensive; thus, future studies may provide novel information regarding the effect of GCs on cell processes. On the other hand, research on GR's role in cancer is quite extensive, but its role still remains elusive. GCs' function in cancer depends on the target cell and differs among various stages of cancer, hence making the study of their role in cancer pathogenesis and pathology extremely complicated.

Apart from publicly available literature data, multiple full-genome datasets from different populations are also publicly available for research. These datasets can help study how genomic alterations in the GR interactome may promote distinct population characteristics. A comparison between Japanese and Korean individuals showcased that most SNPs associated with pathological conditions were present in intronic regions, once again highlighting the importance of introns in biological function. Another observation is that rs1043618 frequency is quite different between the Japanese and Korean individuals, despite mentioned populations being pretty similar on a genetic level. This polymorphism has been associated with COPD in response to environmental stressors in a Mexican population. Since Korean individuals display a higher incidence rate of COPD than Japanese individuals, despite

similar smoking trends, this polymorphism may prove to be of great research interest.

Future research on two specific findings may prove to be quite beneficial. Firstly, the possibility that the Y537S mutation may force ERa to display a structure more similar to that of ERb could be useful in drug design since mentioned mutation has been heavily associated with breast cancer. Identifying antagonists that successfully block this variant of ERa may help advance current therapeutic approaches in breast cancer. The possibility that rs1043618 may be responsible for the disparity in COPD cases between the Japanese and Korean population may help identify individuals with a higher risk of COPD whose monitoring may help disease diagnosis. However, it should be stated that these findings are not definite and further research is needed. Several structures featuring the Y537S mutation also featured other mutations which too may have led to the emergence of the alternate ERa structure. Additionally, despite the large sample size, there exists only one research associating rs1043618 with COPD and the fact that specific SNPs may be associated with a pathological condition in one population but show no association in another one should be taken into consideration.

References

1. Chrousos GP. Stress and disorders of the stress system. Nature Reviews Endocrinology. 2009;5(7):374-81.
2. Nicolaides NC, Kyratzi E, Lamprokostopoulou A, Chrousos GP, Charmandari E. Stress, the stress system and the role of glucocorticoids. Neuroimmunomodulation. 2015;22(1-2):6-19.
3. Charmandari E, Tsigos C, Chrousos G. ENDOCRINOLOGY OF THE STRESS RESPONSE. Annual Review of Physiology. 2005;67(1):259-84.
4. Dunlavey CJ. Introduction to the Hypothalamic-Pituitary-Adrenal Axis: Healthy and Dysregulated Stress Responses, Developmental Stress and Neurodegeneration. J Undergrad Neurosci Educ. 2018;16(2):R59-R60.
5. Chrousos GP, Gold PW. The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. Jama. 1992;267(9):1244-52.
6. McEwen BS, Bowles NP, Gray JD, Hill MN, Hunter RG, Karatsoreos IN, et al. Mechanisms of stress in the brain. Nat Neurosci. 2015;18(10):1353-63.
7. Eisenmann ED, Rorabaugh BR, Zoladz PR. Acute Stress Decreases but Chronic Stress Increases Myocardial Sensitivity to Ischemic Injury in Rodents. Front Psychiatry. 2016;7:71-.
8. Franke HA. Toxic Stress: Effects, Prevention and Treatment. Children (Basel). 2014;1(3):390-402.
9. Benarroch EE. The locus ceruleus norepinephrine system: functional organization and potential clinical significance. Neurology. 2009;73(20):1699-704.
10. Lü Y-F, Yang Y, Li C-L, Wang Y, Li Z, Chen J. The Locus Coeruleus-Norepinephrine System Mediates Empathy for Pain through Selective Up-Regulation of P2X3 Receptor in Dorsal Root Ganglia in Rats. Front Neural Circuits. 2017;11:66-.
11. Kanczkowski W, Sue M, Bornstein SR. Adrenal Gland Microenvironment and Its Involvement in the Regulation of Stress-Induced Hormone Secretion during Sepsis. Front Endocrinol (Lausanne). 2016;7:156-.
12. DeMorrow S. Role of the Hypothalamic-Pituitary-Adrenal Axis in Health and Disease. Int J Mol Sci. 2018;19(4):986.
13. Chrousos GP. Stressors, stress, and neuroendocrine integration of the adaptive response. The 1997 Hans Selye Memorial Lecture. Annals of the New York Academy of Sciences. 1998;851:311-35.
14. Nicolaides NC, Charmandari E, Chrousos GP, Kino T. Circadian endocrine rhythms: the hypothalamic-pituitary-adrenal axis and its actions. Annals of the New York Academy of Sciences. 2014;1318:71-80.
15. Timmermans S, Souffriau J, Libert C. A General Introduction to Glucocorticoid Biology. Frontiers in immunology. 2019;10:1545.
16. Tsigos C, Chrousos GP. Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. Journal of Psychosomatic Research. 2002;53(4):865-71.
17. Stokes J, Noble J, Brett L, Phillips C, Seckl JR, O'Brien C, et al. Distribution of glucocorticoid and mineralocorticoid receptors and 11beta-hydroxysteroid dehydrogenases in human and rat ocular tissues. Investigative ophthalmology & visual science. 2000;41(7):1629-38.
18. Sevilla LM, Pérez P. Roles of the Glucocorticoid and Mineralocorticoid Receptors in Skin Pathophysiology. Int J Mol Sci. 2018;19(7):1906.
19. Viengchareun S, Le Menuet D, Martinerie L, Munier M, Pascual-Le Tallec L, Lombès M. The mineralocorticoid receptor: insights into its molecular and (patho)physiological biology. Nucl Recept Signal. 2007;5:e012-e.

20. Koning A, Buurstede JC, van Weert L, Meijer OC. Glucocorticoid and Mineralocorticoid Receptors in the Brain: A Transcriptional Perspective. Journal of the Endocrine Society. 2019;3(10):1917-30.
21. Kumar R, Litwack G. Structural and functional relationships of the steroid hormone receptors' N-terminal transactivation domain. Steroids. 2009;74(12):877-83.
22. Tao LJ, Seo DE, Jackson B, Ivanova NB, Santori FR. Nuclear Hormone Receptors and Their Ligands: Metabolites in Control of Transcription. Cells. 2020;9(12):2606.
23. Rastinejad F, Huang P, Chandra V, Khorasanizadeh S. Understanding nuclear receptor form and function using structural biology. J Mol Endocrinol. 2013;51(3):T1-T21.
24. Porter BA, Ortiz MA, Bratslavsky G, Kotula L. Structure and Function of the Nuclear Receptor Superfamily and Current Targeted Therapies of Prostate Cancer. Cancers (Basel). 2019;11(12):1852.
25. Pawlak M, Lefebvre P, Staels B. General molecular biology and architecture of nuclear receptors. Curr Top Med Chem. 2012;12(6):486-504.
26. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: A structural perspective. Protein Science. 2018;27(11):1876-92.
27. Helsen C, Kerkhofs S, Clinckemalie L, Spans L, Laurent M, Boonen S, et al. Structural basis for nuclear hormone receptor DNA binding. Molecular and cellular endocrinology. 2012;348(2):411-7.
28. Folkertsma S, van Noort PI, Brandt RF, Bettler E, Vriend G, de Vlieg J. The nuclear receptor ligand-binding domain: a family-based structure analysis. Current medicinal chemistry. 2005;12(9):1001-16.
29. Bain DL, Heneghan AF, Connaghan-Jones KD, Miura MT. Nuclear receptor structure: implications for function. Annu Rev Physiol. 2007;69:201-20.
30. Batista MRB, Martínez L. Conformational Diversity of the Helix 12 of the Ligand Binding Domain of PPARγ and Functional Implications. The Journal of Physical Chemistry B. 2015;119(50):15418-29.
31. McGrady AV. Effects of psychological stress on male reproduction: a review. Archives of andrology. 1984;13(1):1-7.
32. Cunningham RL, Lumia AR, McGinnis MY. Androgen receptors, sex behavior, and aggression. Neuroendocrinology. 2012;96(2):131-40.
33. Handa RJ, Mani SK, Uht RM. Estrogen receptors and the regulation of neural stress responses. Neuroendocrinology. 2012;96(2):111-8.
34. Ulrich-Lai YM, Ryan KK. PPARγ and stress: implications for aging. Exp Gerontol. 2013;48(7):671-6.
35. Buchanan TW, Lovallo WR. The role of genetics in stress effects on health and addiction. Curr Opin Psychol. 2019;27:72-6.
36. Ising M, Holsboer F. Genetics of stress response and stress-related disorders. Dialogues Clin Neurosci. 2006;8(4):433-44.
37. Pagliaccio D, Luby JL, Bogdan R, Agrawal A, Gaffrey MS, Belden AC, et al. Stress-system genes and life stress predict cortisol levels and amygdala and hippocampal volumes in children. Neuropsychopharmacology. 2014;39(5):1245-53.
38. Bogdan R, Santesso DL, Fagerness J, Perlis RH, Pizzagalli DA. Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011;31(37):13246-54.
39. Bet PM, Penninx BW, Bochdanovits Z, Uitterlinden AG, Beekman AT, van Schoor NM, et al. Glucocorticoid receptor gene polymorphisms and childhood adversity are associated with depression: New evidence for a gene-environment interaction. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2009;150b(5):660-9.

40. Mihaljevic M, Zeljic K, Soldatovic I, Andric S, Mirjanic T, Richards A, et al. The emerging role of the FKBP5 gene polymorphisms in vulnerability-stress model of schizophrenia: further evidence from a Serbian population. European archives of psychiatry and clinical neuroscience. 2017;267(6):527-39.
41. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. Nature. 2019;571(7766):489-99.
42. Mazzio EA, Soliman KF. Basic concepts of epigenetics: impact of environmental signals on gene expression. Epigenetics. 2012;7(2):119-30.
43. Sarkies P. Molecular mechanisms of epigenetic inheritance: Possible evolutionary implications. Semin Cell Dev Biol. 2020;97:106-15.
44. Matosin N, Cruceanu C, Binder EB. Preclinical and Clinical Evidence of DNA Methylation Changes in Response to Trauma and Chronic Stress. Chronic stress (Thousand Oaks, Calif). 2017;1.
45. Jiang S, Postovit L, Cattaneo A, Binder EB, Aitchison KJ. Epigenetic Modifications in Stress Response Genes Associated With Childhood Trauma. Front Psychiatry. 2019;10:808-.
46. Hing B, Gardner C, Potash JB. Effects of negative stressors on DNA methylation in the brain: implications for mood and anxiety disorders. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2014;165B(7):541-54.
47. Seematter G, Binnert C, Martin JL, Tappy L. Relationship between stress, inflammation and metabolism. Current opinion in clinical nutrition and metabolic care. 2004;7(2):169-73.
48. Vegiopoulos A, Herzig S. Glucocorticoids, metabolism and metabolic diseases. Molecular and cellular endocrinology. 2007;275(1-2):43-61.
49. Konstandi M, Johnson EO, Lang MA. Consequences of psychophysiological stress on cytochrome P450-catalyzed drug metabolism. Neuroscience and biobehavioral reviews. 2014;45:149-67.
50. Manikandan P, Nagini S. Cytochrome P450 Structure, Function and Clinical Significance: A Review. Current drug targets. 2018;19(1):38-54.
51. Schiffer L, Barnard L, Baranowski ES, Gilligan LC, Taylor AE, Arlt W, et al. Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. J Steroid Biochem Mol Biol. 2019;194:105439-.
52. Custodio JM, Donaldson KM, Hunt HJ. An In Vitro and In Vivo Evaluation of the Effect of Relacorilant on the Activity of Cytochrome P450 Drug Metabolizing Enzymes. Journal of clinical pharmacology. 2021;61(2):244-53.
53. Pantelidou M, Tsiakitzis K, Rekka EA, Kourounakis PN. Biologic Stress, Oxidative Stress, and Resistance to Drugs: What Is Hidden Behind. Molecules. 2017;22(2):307.
54. Tsigos C, Kyrou I, Kassi E, Chrousos GP. Stress: Endocrine Physiology and Pathophysiology Endotext [Internet]: 2020 Oct 17; 2000 [updated 2020 Oct 17.
55. Kalantaridou SN, Makrigiannakis A, Zoumakis E, Chrousos GP. Stress and the female reproductive system. Journal of reproductive immunology. 2004;62(1-2):61-8.
56. Dhabhar FS. Effects of stress on immune function: the good, the bad, and the beautiful. Immunologic research. 2014;58(2-3):193-210.
57. Morey JN, Boggero IA, Scott AB, Segerstrom SC. Current Directions in Stress and Human Immune Function. Curr Opin Psychol. 2015;5:13-7.
58. Cain DW, Cidlowski JA. Immune regulation by glucocorticoids. Nature Reviews Immunology. 2017;17(4):233-47.
59. Ronchetti S, Migliorati G, Bruscoli S, Riccardi C. Defining the role of glucocorticoids in inflammation. Clinical science (London, England : 1979). 2018;132(14):1529-43.

60. McEwen BS, Gianaros PJ. Central role of the brain in stress and adaptation: links to socioeconomic status, health, and disease. Annals of the New York Academy of Sciences. 2010;1186:190-222.
61. McEwen BS. Protective and damaging effects of stress mediators: central role of the brain. Dialogues Clin Neurosci. 2006;8(4):367-81.
62. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: A review. EXCLI J. 2017;16:1057-72.
63. Vyas S, Rodrigues AJ, Silva JM, Tronche F, Almeida OFX, Sousa N, et al. Chronic Stress and Glucocorticoids: From Neuronal Plasticity to Neurodegeneration. Neural Plast. 2016;2016:6391686-.
64. Canet G, Chevallier N, Zussy C, Desrumaux C, Givalois L. Central Role of Glucocorticoid Receptors in Alzheimer's Disease and Depression. Front Neurosci. 2018;12:739-.
65. Mariotti A. The effects of chronic stress on health: new insights into the molecular mechanisms of brain-body communication. Future Sci OA. 2015;1(3):FSO23-FSO.
66. McEwen BS, Nasca C, Gray JD. Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex. Neuropsychopharmacology. 2016;41(1):3-23.
67. Lisman J, Buzsáki G, Eichenbaum H, Nadel L, Ranganath C, Redish AD. Viewpoints: how the hippocampus contributes to memory, navigation and cognition. Nat Neurosci. 2017;20(11):1434-47.
68. Kim EJ, Pellman B, Kim JJ. Stress effects on the hippocampus: a critical review. Learn Mem. 2015;22(9):411-6.
69. Vyas A, Bernal S, Chattarji S. Effects of chronic stress on dendritic arborization in the central and extended amygdala. Brain research. 2003;965(1-2):290-4.
70. Vyas A, Mitra R, Shankaranarayana Rao BS, Chattarji S. Chronic stress induces contrasting patterns of dendritic remodeling in hippocampal and amygdaloid neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2002;22(15):6810-8.
71. Guadagno A, Belliveau C, Mechawar N, Walker C-D. Effects of Early Life Stress on the Developing Basolateral Amygdala-Prefrontal Cortex Circuit: The Emerging Role of Local Inhibition and Perineuronal Nets. Frontiers in Human Neuroscience. 2021;15(484).
72. Hölzel BK, Carmody J, Evans KC, Hoge EA, Dusek JA, Morgan L, et al. Stress reduction correlates with structural changes in the amygdala. Soc Cogn Affect Neurosci. 2010;5(1):11-7.
73. Arnsten AFT. Stress signalling pathways that impair prefrontal cortex structure and function. Nat Rev Neurosci. 2009;10(6):410-22.
74. Baumeister D, Lightman SL, Pariante CM. The Interface of Stress and the HPA Axis in Behavioural Phenotypes of Mental Illness. Current topics in behavioral neurosciences. 2014;18:13-24.
75. Menke A. Is the HPA Axis as Target for Depression Outdated, or Is There a New Hope? Front Psychiatry. 2019;10:101-.
76. Yang L, Zhao Y, Wang Y, Liu L, Zhang X, Li B, et al. The Effects of Psychological Stress on Depression. Curr Neuropharmacol. 2015;13(4):494-504.
77. Raghupathi W, Raghupathi V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach. Int J Environ Res Public Health. 2018;15(3):431.
78. Volden PA, Conzen SD. The influence of glucocorticoid signaling on tumor progression. Brain Behav Immun. 2013;30 Suppl(0):S26-S31.
79. Lin K-T, Wang L-H. New dimension of glucocorticoids in cancer treatment. Steroids. 2016;111:84-8.
80. Dai S, Mo Y, Wang Y, Xiang B, Liao Q, Zhou M, et al. Chronic Stress Promotes Cancer Development. Front Oncol. 2020;10:1492-.

81. Gupta A, Shah K, Oza MJ, Behl T. Reactivation of p53 gene by MDM2 inhibitors: A novel therapy for cancer treatment. Biomedicine & Pharmacotherapy. 2019;109:484-92.
82. Steptoe A, Kivimäki M. Stress and cardiovascular disease. Nature reviews Cardiology. 2012;9(6):360-70.
83. Roemmich JN, Lambiase MJ, Balantekin KN, Feda DM, Dorn J. Stress, behavior, and biology: risk factors for cardiovascular diseases in youth. Exerc Sport Sci Rev. 2014;42(4):145-52.
84. Gordan R, Gwathmey JK, Xie L-H. Autonomic and endocrine control of cardiovascular function. World J Cardiol. 2015;7(4):204-14.
85. Momen A, Mascarenhas V, Gahremanpour A, Gao Z, Moradkhan R, Kunselman A, et al. Coronary blood flow responses to physiological stress in humans. Am J Physiol Heart Circ Physiol. 2009;296(3):H854-H61.
86. Wei P, Li Y, Wu L, Wu J, Wu W, Chen S, et al. Serum cortisol levels and adrenal gland size in patients with chronic obstructive pulmonary disease. Am J Transl Res. 2021;13(7):8150-7.
87. Newsome BR, McDonnell K, Hucks J, Dawson Estrada R. Chronic Obstructive Pulmonary Disease: Clinical Implications for Patients With Lung Cancer. Clin J Oncol Nurs. 2018;22(2):184-92.
88. Pembroke TPI, Rasul F, Hart CL, Davey Smith G, Stansfeld SA. Psychological distress and chronic obstructive pulmonary disease in the Renfrew and Paisley (MIDSPAN) study. J Epidemiol Community Health. 2006;60(9):789-92.
89. Kelly SJ, Ismail M. Stress and type 2 diabetes: a review of how stress contributes to the development of type 2 diabetes. Annual review of public health. 2015;36:441-62.
90. Lloyd C, Smith J, Weinger K. Stress and Diabetes: A Review of the Links. Diabetes Spectrum. 2005;18(2):121-7.
91. Hackett RA, Steptoe A. Type 2 diabetes mellitus and psychological stress — a modifiable risk factor. Nature Reviews Endocrinology. 2017;13(9):547-60.
92. Kadmiel M, Cidlowski JA. Glucocorticoid receptor signaling in health and disease. Trends Pharmacol Sci. 2013;34(9):518-30.
93. Tan CK, Wahli W. A trilogy of glucocorticoid receptor actions. Proc Natl Acad Sci U S A. 2016;113(5):1115-7.
94. Kaziales A, Barkovits K, Marcus K, Richter K. Glucocorticoid receptor complexes form cooperatively with the Hsp90 co-chaperones Pp5 and FKBPs. Sci Rep. 2020;10(1):10733-.
95. Sinclair D, Fillman SG, Webster MJ, Weickert CS. Dysregulation of glucocorticoid receptor co-factors FKBP5, BAG1 and PTGES3 in prefrontal cortex in psychotic illness. Sci Rep. 2013;3(1):3539.
96. Baker JD, Ozsan I, Rodriguez Ospina S, Gulick D, Blair LJ. Hsp90 Heterocomplexes Regulate Steroid Hormone Receptors: From Stress Response to Psychiatric Disease. Int J Mol Sci. 2018;20(1):79.
97. Louw A. GR Dimerization and the Impact of GR Dimerization on GR Protein Stability and Half-Life. Frontiers in immunology. 2019;10:1693-.
98. Robertson S, Hapgood JP, Louw A. Glucocorticoid receptor concentration and the ability to dimerize influence nuclear translocation and distribution. Steroids. 2013;78(2):182-94.
99. Frego L, Davidson W. Conformational changes of the glucocorticoid receptor ligand binding domain induced by ligand and cofactor binding, and the location of cofactor binding sites determined by hydrogen/deuterium exchange mass spectrometry. Protein Sci. 2006;15(4):722-30.
100. Vandevyver S, Dejager L, Libert C. On the trail of the glucocorticoid receptor: into the nucleus and back. Traffic (Copenhagen, Denmark). 2012;13(3):364-74.

101. Hudson WH, Youn C, Ortlund EA. The structural basis of direct glucocorticoid-mediated transrepression. Nat Struct Mol Biol. 2013;20(1):53-8.
102. Groeneweg FL, van Royen ME, Fenz S, Keizer VIP, Geverts B, Prins J, et al. Quantitation of glucocorticoid receptor DNA-binding dynamics by single-molecule microscopy and FRAP. PLoS One. 2014;9(3):e90532-e.
103. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. J Integr Bioinform. 2018;15(3):20170030.
104. Mallappallil M, Sabu J, Gruessner A, Salifu M. A review of big data and medical research. SAGE Open Med. 2020;8:2050312120934839-.
105. Galetsi P, Katsaliaki K, Kumar S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. Social science & medicine (1982). 2019;241:112533.
106. Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. BMC Med. 2019;17(1):143-.
107. Benke K, Benke G. Artificial Intelligence and Big Data in Public Health. Int J Environ Res Public Health. 2018;15(12):2796.
108. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From Big Data to Precision Medicine. Front Med (Lausanne). 2019;6:34-.
109. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics. 2008;9(5):356-69.
110. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Communications Biology. 2019;2(1):9.
111. Chang M, He L, Cai L. An Overview of Genome-Wide Association Studies. Methods in molecular biology (Clifton, NJ). 2018;1754:97-108.
112. Liang B, Ding H, Huang L, Luo H, Zhu X. GWAS in cancer: progress and challenges. Molecular Genetics and Genomics. 2020;295(3):537-61.
113. Wong K-C. Big data challenges in genome informatics. Biophys Rev. 2019;11(1):51-4.
114. Ginsburg GS, Phillips KA. Precision Medicine: From Science To Value. Health Aff (Millwood). 2018;37(5):694-701.
115. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertil Steril. 2018;109(6):952-63.
116. Abrahams E. Right drug-right patient-right time: personalized medicine coalition. Clin Transl Sci. 2008;1(1):11-2.
117. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Research. 2000;28(1):235-42.
118. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research. 2017;45(D1):D896-D901.
119. Emerson JJ, Li W-H. The genetic basis of evolutionary change in gene expression levels. Philos Trans R Soc Lond B Biol Sci. 2010;365(1552):2581-90.
120. Breidbach O, Ghiselin MT. Evolution and development: Past, present, and future. Theory in Biosciences. 2007;125(2):157-71.
121. Reilly SK, Noonan JP. Evolution of Gene Regulation in Humans. Annual review of genomics and human genetics. 2016;17:45-67.
122. García-Sánchez A, Marqués-García F. Review of Methods to Study Gene Expression Regulation Applied to Asthma. Methods in molecular biology (Clifton, NJ). 2016;1434:71-89.
123. Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity. 2010;105(1):4-13.
124. Atkinson TJ, Halfon MS. Regulation of gene expression in the genomic context. Comput Struct Biotechnol J. 2014;9:e201401001-e.

125. Day DA, Tuite MF. Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. The Journal of endocrinology. 1998;157(3):361-71.
126. Laker RC, Ryall JG. DNA Methylation in Skeletal Muscle Stem Cell Specification, Proliferation, and Differentiation. Stem cells international. 2016;2016:5725927.
127. Alaskhar Alhamwe B, Khalaila R, Wolf J, von Bülow V, Harb H, Alhamdan F, et al. Histone modifications and their role in epigenetics of atopy and allergic diseases. Allergy Asthma Clin Immunol. 2018;14:39-.
128. Lorch Y, Maier-Davis B, Kornberg RD. Mechanism of chromatin remodeling. Proc Natl Acad Sci U S A. 2010;107(8):3458-62.
129. Duan G, Walther D. The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput Biol. 2015;11(2):e1004049-e.
130. Cai Y, Yu X, Hu S, Yu J. A brief review on the mechanisms of miRNA regulation. Genomics Proteomics Bioinformatics. 2009;7(4):147-54.
131. Latchman DS. Transcription factors: an overview. Int J Exp Pathol. 1993;74(5):417-22.
132. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172(4):650-65.
133. Klenk H-P, Doolittle WF. Evolution: Archaea and eukaryotes versus bacteria? Current Biology. 1994;4(10):920-2.
134. Lee DJ, Minchin SD, Busby SJ. Activating transcription in bacteria. Annual review of microbiology. 2012;66:125-52.
135. Venters BJ, Pugh BF. How eukaryotic genes are transcribed. Critical reviews in biochemistry and molecular biology. 2009;44(2-3):117-41.
136. Gehring AM, Walker JE, Santangelo TJ. Transcription Regulation in Archaea. J Bacteriol. 2016;198(14):1906-17.
137. Adcock IM, Caramori G. Chapter 31 - Transcription Factors. In: Barnes PJ, Drazen JM, Rennard SI, Thomson NC, editors. Asthma and COPD (Second Edition). Oxford: Academic Press; 2009. p. 373-80.
138. Powell RV, Willett CR, Goertzen LR, Rashotte AM. Lineage specific conservation of cis-regulatory elements in Cytokinin Response Factors. Sci Rep. 2019;9(1):13387.
139. Wilkinson AC, Nakauchi H, Göttgens B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. Cell systems. 2017;5(4):319-31.
140. Rebeiz M, Tsiantis M. Enhancer evolution and the origins of morphological novelty. Current opinion in genetics & development. 2017;45:115-23.
141. Yesudhas D, Batool M, Anwar MA, Panneerselvam S, Choi S. Proteins Recognizing DNA: Structural Uniqueness and Versatility of DNA-Binding Domains in Stem Cell Transcription Factors. Genes. 2017;8(8).
142. Cheatle Jarvela AM, Hinman VF. Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. EvoDevo. 2015;6(1):3.
143. Zhaxybayeva O, Lapierre P, Gogarten JP. Ancient gene duplications and the root(s) of the tree of life. Protoplasma. 2005;227(1):53-64.
144. Albalat R, Cañestro C. Evolution by gene loss. Nature reviews Genetics. 2016;17(7):379-91.
145. Rosanova A, Colliva A, Osella M, Caselle M. Modelling the evolution of transcription factor binding preferences in complex eukaryotes. Sci Rep. 2017;7(1):7596.
146. Pérez-Rueda E, Collado-Vides J, Segovia L. Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea. Computational biology and chemistry. 2004;28(5-6):341-50.
147. Duttke SH. Evolution and diversification of the basal transcription machinery. Trends in biochemical sciences. 2015;40(3):127-9.

148. Duttke SHC. Evolution and diversification of the basal transcription machinery. Trends in biochemical sciences. 2015;40(3):127-9.
149. Charoensawan V, Wilson D, Teichmann SA. Genomic repertoires of DNA-binding transcription factors across the tree of life. Nucleic acids research. 2010;38(21):7364-77.
150. Inukai S, Kock KH, Bulyk ML. Transcription factor-DNA binding: beyond binding site motifs. Current opinion in genetics & development. 2017;43:110-9.
151. Siepel A, Arbiza L. Cis-regulatory elements and human evolution. Current opinion in genetics & development. 2014;29:81-9.
152. Tuğrul M, Paixão T, Barton NH, Tkačik G. Dynamics of Transcription Factor Binding Site Evolution. PLoS genetics. 2015;11(11):e1005639.
153. Berg J, Willmann S, Lässig M. Adaptive evolution of transcription factor binding sites. BMC evolutionary biology. 2004;4:42.
154. Yang S, Yalamanchili HK, Li X, Yao KM, Sham PC, Zhang MQ, et al. Correlated evolution of transcription factors and their binding sites. Bioinformatics (Oxford, England). 2011;27(21):2972-8.
155. Morgunova E, Taipale J. Structural perspective of cooperative transcription factor binding. Current opinion in structural biology. 2017;47:1-8.
156. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. Trends in biochemical sciences. 2008;33(5):220-9.
157. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 2016;167(5):1170-87.
158. Miller JA, Widom J. Collaborative competition mechanism for gene activation in vivo. Molecular and cellular biology. 2003;23(5):1623-32.
159. Rosenfeld MG, Lunyak VV, Glass CK. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes & development. 2006;20(11):1405-28.
160. Everett L, Hansen M, Hannenhalli S. Regulating the regulators: modulators of transcription factor activity. Methods in molecular biology (Clifton, NJ). 2010;674:297-312.
161. Tootle TL, Rebay I. Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. BioEssays : news and reviews in molecular, cellular and developmental biology. 2005;27(3):285-98.
162. Reiter F, Wienerroither S, Stark A. Combinatorial function of transcription factors and cofactors. Current opinion in genetics & development. 2017;43:73-81.
163. Lynch VJ, May G, Wagner GP. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature. 2011;480(7377):383-6.
164. Lynch VJ, Wagner GP. Resurrecting the role of transcription factor change in developmental evolution. Evolution; international journal of organic evolution. 2008;62(9):2131-54.
165. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nature reviews Genetics. 2009;10(4):252-63.
166. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends in genetics : TIG. 2016;32(1):42-56.
167. Xin B, Rohs R. Relationship between histone modifications and transcription factor binding is protein family specific. Genome research. 2018;28(3):321-33.
168. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Frontiers in Endocrinology. 2018;9(402).
169. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews Genetics. 2007;8(2):93-103.

170. Qiu C, Wang J, Yao P, Wang E, Cui Q. microRNA evolution in a human transcription factor and microRNA regulatory network. BMC Syst Biol. 2010;4:90-.
171. Holzer G, Markov GV, Laudet V. Evolution of Nuclear Receptors and Ligand Signaling: Toward a Soft Key-Lock Model? Current topics in developmental biology. 2017;125:1-38.
172. Markov GV, Laudet V. Origin and evolution of the ligand-binding ability of nuclear receptors. Molecular and cellular endocrinology. 2011;334(1-2):21-30.
173. MATLAB. R2018a. The MathWorks Inc.; 2018.
174. Papageorgiou L, Loukatou S, Sofia K, Maroulis D, Vlachakis D. An updated evolutionary study of Flaviviridae NS3 helicase and NS5 RNA-dependent RNA polymerase reveals novel invariable motifs as potential pharmacological targets. Molecular BioSystems. 2016;12(7):2080-93.
175. Pearson WR. Selecting the Right Similarity-Scoring Matrix. Current protocols in bioinformatics. 2013;43:3.5.1-3.5.9.
176. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics (Oxford, England). 2009;25(9):1189-91.
177. MOE. Molecular Operating Environment. Chemical Computing Group ULC: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada; 2014.1.
178. Kufareva I, Abagyan R. Methods of protein structure comparison. Methods in molecular biology (Clifton, NJ). 2012;857:231-57.
179. Aertgeerts K, Skene R, Yano J, Sang BC, Zou H, Snell G, et al. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. The Journal of biological chemistry. 2011;286(21):18756-65.
180. Papageorgiou L, Loukatou S, Koumandou VL, Makałowski W, Megalooikonomou V, Vlachakis D, et al. Structural models for the design of novel antiviral agents against Greek Goat Encephalitis. PeerJ. 2014;2:e664-e.
181. Papageorgiou L, Megalooikonomou V, Vlachakis D. Genetic and structural study of DNA-directed RNA polymerase II of Trypanosoma brucei, towards the designing of novel antiparasitic agents. PeerJ. 2017;5:e3061.
182. Michener CD, Sokal RR. A QUANTITATIVE APPROACH TO A PROBLEM IN CLASSIFICATION. Evolution; international journal of organic evolution. 1957;11(2):130-62.
183. Sneath PHA, RR S. Unweighted pair group method with arithmetic mean. Numerical Taxonomy. 1973:pp230‐4.
184. Lu J, Xu G, Zhang S, Lu B. An effective sequence-alignment-free superpositioning of pairwise or multiple structures with missing data. Algorithms for Molecular Biology. 2016;11(1):18.
185. Pavlopoulos GA, Soldatos TG, Barbosa-Silva A, Schneider R. A reference guide for tree analysis and visualization. BioData mining. 2010;3(1):1.
186. Leaché AD, Wagner P, Linkem CW, Böhme W, Papenfuss TJ, Chong RA, et al. A hybrid phylogenetic–phylogenomic approach for species tree estimation in African Agama lizards with applications to biogeography, character evolution, and diversification. Molecular Phylogenetics and Evolution. 2014;79:215-30.
187. Fouquier J, Rideout JR, Bolyen E, Chase J, Shiffer A, McDonald D, et al. Ghost-tree: creating hybrid-gene phylogenetic trees for diversity analyses. Microbiome. 2016;4:11.
188. Stecher G, Liu L, Sanderford M, Peterson D, Tamura K, Kumar S. MEGA-MD: molecular evolutionary genetics analysis software with mutational diagnosis of amino acid variation. Bioinformatics (Oxford, England). 2014;30(9):1305-7.
189. Mellor CL, Marchese Robinson RL, Benigni R, Ebbrell D, Enoch SJ, Firman JW, et al. Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use. Regulatory toxicology and pharmacology : RTP. 2019;101:121-34.

190. Rácz A, Bajusz D, Héberger K. Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. Journal of cheminformatics. 2018;10(1):48.
191. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of cheminformatics. 2015;7(1):20.
192. Fayard E, Auwerx J, Schoonjans K. LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. Trends in cell biology. 2004;14(5):250-60.
193. Ghbeish N, Tsai C-C, Schubiger M, Zhou JY, Evans RM, McKeown M. The dual role of ultraspiracle, the Drosophila retinoid X receptor, in the ecdysone response. Proceedings of the National Academy of Sciences. 2001;98(7):3867-72.
194. Laffitte BA, Kast HR, Nguyen CM, Zavacki AM, Moore DD, Edwards PA. Identification of the DNA binding specificity and potential target genes for the farnesoid X-activated receptor. The Journal of biological chemistry. 2000;275(14):10638-47.
195. Son YL, Park MJ, Lee YC. General and specific determinants of the selective interactions between SRC-1 NR box-2 and target nuclear receptors. Molecular biology reports. 2012;39(1):177-84.
196. Johansson L, Båvner A, Thomsen JS, Färnegårdh M, Gustafsson JA, Treuter E. The orphan nuclear receptor SHP utilizes conserved LXXLL-related motifs for interactions with ligand-activated estrogen receptors. Molecular and cellular biology. 2000;20(4):1124-33.
197. Bernardo TJ, Dubrovsky EB. The Drosophila juvenile hormone receptor candidates methoprene-tolerant (MET) and germ cell-expressed (GCE) utilize a conserved LIXXL motif to bind the FTZ-F1 nuclear receptor. The Journal of biological chemistry. 2012;287(10):7821-33.
198. Nicolaides NC, Roberts ML, Kino T, Braatvedt G, Hurt DE, Katsantoni E, et al. A novel point mutation of the human glucocorticoid receptor gene causes primary generalized glucocorticoid resistance through impaired interaction with the LXXLL motif of the p160 coactivators: dissociation of the transactivating and transreppressive activities. The Journal of clinical endocrinology and metabolism. 2014;99(5):E902-7.
199. Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, McKee DD, et al. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. Cell. 2002;110(1):93-105.
200. Jääskeläinen J, Mongan NP, Harland S, Hughes IA. Five novel androgen receptor gene mutations associated with complete androgen insensitivity syndrome. Human mutation. 2006;27(3):291.
201. Vitellius G, Fagart J, Delemer B, Amazit L, Ramos N, Bouligand J, et al. Three Novel Heterozygous Point Mutations of NR3C1 Causing Glucocorticoid Resistance. Human mutation. 2016;37(8):794-803.
202. Harrod A, Fulton J, Nguyen VTM, Periyasamy M, Ramos-Garcia L, Lai CF, et al. Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. Oncogene. 2017;36(16):2286-96.
203. Achermann JC, Schwabe J, Fairall L, Chatterjee K. Genetic disorders of nuclear receptors. The Journal of clinical investigation. 2017;127(4):1181-92.
204. Nicolaides NC, Charmandari E. Chrousos syndrome: from molecular pathogenesis to therapeutic management. European Journal of Clinical Investigation. 2015;45(5):504-14.
205. Ai N, Krasowski MD, Welsh WJ, Ekins S. Understanding nuclear receptors using computational methods. Drug discovery today. 2009;14(9-10):486-94.
206. Matias PM, Carrondo MA, Coelho R, Thomaz M, Zhao XY, Wegg A, et al. Structural basis for the glucocorticoid response in a mutant human androgen receptor (AR(ccr)) derived from an androgen-independent prostate cancer. Journal of medicinal chemistry. 2002;45(7):1439-46.
207. Puyang X, Furman C, Zheng GZ, Wu ZJ, Banka D, Aithal K, et al. Discovery of Selective Estrogen Receptor Covalent Antagonists for the Treatment of ERα(WT) and ERα(MUT) Breast Cancer. Cancer discovery. 2018;8(9):1176-93.

208. Nettles KW, Bruning JB, Gil G, O'Neill EE, Nowak J, Guo Y, et al. Structural plasticity in the oestrogen receptor ligand-binding domain. EMBO reports. 2007;8(6):563-8.
209. Siltberg-Liberles J, Grahnen JA, Liberles DA. The evolution of protein structures and structural ensembles under functional constraint. Genes. 2011;2(4):748-62.
210. Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, Moras D, et al. A canonical structure for the ligand-binding domain of nuclear receptors. Nature structural biology. 1996;3(1):87-94.
211. Papageorgiou L, Zervou MI, Vlachakis D, Matalliotakis M, Matalliotakis I, Spandidos DA, et al. Demetra Application: An integrated genotype analysis web server for clinical genomics in endometriosis. International journal of molecular medicine. 2021;47(6).
212. Ossom Williamson P, Minter CIJ. Exploring PubMed as a reliable resource for scholarly communications services. J Med Libr Assoc. 2019;107(1):16-29.
213. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.
214. Allot A, Peng Y, Wei C-H, Lee K, Phan L, Lu Z. LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. Nucleic acids research. 2018;46(W1):W530-W6.
215. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research. 2014;42(Database issue):D980-D5.
216. Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. 2007;35(4):495-516.
217. Gruver-Yates AL, Cidlowski JA. Tissue-specific actions of glucocorticoids on apoptosis: a double-edged sword. Cells. 2013;2(2):202-23.
218. Freishtat RJ, Nagaraju K, Jusko W, Hoffman EP. Glucocorticoid efficacy in asthma: is improved tissue remodeling upstream of anti-inflammation. J Investig Med. 2010;58(1):19-22.
219. Montgomery B, Cheng HH, Drechsler J, Mostaghel EA. Glucocorticoids and prostate cancer treatment: friend or foe? Asian J Androl. 2014;16(3):354-8.
220. Valle S, Sharifi N. Targeting Glucocorticoid Metabolism in Prostate Cancer. Endocrinology. 2021;162(9).
221. Rudrapal M, Khairnar SJ, Borse LB, Jadhav AG. Coronavirus Disease-2019 (COVID-19): An Updated Review. Drug research. 2020;70(9):389-400.
222. Alexaki VI, Henneicke H. The Role of Glucocorticoids in the Management of COVID-19. Horm Metab Res. 2021;53(1):9-15.
223. Isidori AM, Arnaldi G, Boscaro M, Falorni A, Giordano C, Giordano R, et al. COVID-19 infection and glucocorticoids: update from the Italian Society of Endocrinology Expert Opinion on steroid replacement in adrenal insufficiency. J Endocrinol Invest. 2020;43(8):1141-7.
224. Briot K, Roux C. Glucocorticoid-induced osteoporosis. RMD Open. 2015;1(1):e000014-e.
225. Strehl C, Ehlers L, Gaber T, Buttgereit F. Glucocorticoids-All-Rounders Tackling the Versatile Players of the Immune System. Frontiers in immunology. 2019;10:1744-.
226. Pufall MA. Glucocorticoids and Cancer. Adv Exp Med Biol. 2015;872:315-33.
227. Guo S, Dipietro LA. Factors affecting wound healing. J Dent Res. 2010;89(3):219-29.
228. Patil VS, Zhou R, Rana TM. Gene regulation by non-coding RNAs. Critical reviews in biochemistry and molecular biology. 2014;49(1):16-32.
229. Lee N, Steitz JA. Noncoding RNA-guided recruitment of transcription factors: A prevalent but undocumented mechanism? BioEssays : news and reviews in molecular, cellular and developmental biology. 2015;37(9):936-41.

230. Shibuya M. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. Genes Cancer. 2011;2(12):1097-105.
231. Noack Watt KE, Achilleos A, Neben CL, Merrill AE, Trainor PA. The Roles of RNA Polymerase I and III Subunits Polr1c and Polr1d in Craniofacial Development and in Zebrafish Models of Treacher Collins Syndrome. PLoS genetics. 2016;12(7):e1006187-e.
232. Liu B, Goodwin JE. The Effect of Glucocorticoids on Angiogenesis in the Treatment of Solid Tumors. J Cell Signal. 2020;1(3):42-9.
233. Frey A, Seifart KH. Glucocorticoids directly affect the synthesis of ribosomal RNA in rat-liver cells. Molecular and cellular endocrinology. 1982;28(2):161-72.
234. Matsui H, Yazawa H, Suzuki N, Hosoya T. Effects of glucocorticoid and cycloheximide on the activity and amount of RNA polymerase I in nuclei of rat liver. The Biochemical journal. 1986;235(3):699-705.
235. Baruteau J, Broomfield A, Crook V, Finnegan N, Harvey K, Burke D, et al. Successful Desensitisation in a Patient with CRIM-Positive Infantile-Onset Pompe Disease. JIMD Rep. 2014;12:99-102.
236. Di Rocco M, Buzzi D, Tarò M. Glycogen storage disease type II: clinical overview. Acta Myol. 2007;26(1):42-4.
237. Dornelles AD, Junges APP, Pereira TV, Krug BC, Gonçalves CBT, Llerena JC, Jr., et al. A Systematic Review and Meta-Analysis of Enzyme Replacement Therapy in Late-Onset Pompe Disease. J Clin Med. 2021;10(21):4828.
238. Benedict M, Zhang X. Non-alcoholic fatty liver disease: An expanded review. World J Hepatol. 2017;9(16):715-32.
239. Woods CP, Hazlehurst JM, Tomlinson JW. Glucocorticoids and non-alcoholic fatty liver disease. J Steroid Biochem Mol Biol. 2015;154:94-103.
240. Jo B-S, Choi SS. Introns: The Functional Benefits of Introns in Genomes. Genomics Inform. 2015;13(4):112-8.
241. Bosia C, Osella M, Baroudi ME, Corà D, Caselle M. Gene autoregulation via intronic microRNAs and its functions. BMC Syst Biol. 2012;6:131.
242. Zhao Z, Dammert MA, Hoppe S, Bierhoff H, Grummt I. Heat shock represses rRNA synthesis by inactivation of TIF-IA and lncRNA-dependent changes in nucleosome positioning. Nucleic acids research. 2016;44(17):8144-52.
243. Boulon S, Westman BJ, Hutten S, Boisvert F-M, Lamond AI. The Nucleolus under Stress. Molecular Cell. 2010;40(2):216-27.
244. Mahajan PB, Thompson EA, Jr. Copurification of RNA polymerase I and the glucocorticoid-regulated transcription factor IC. Protein expression and purification. 1992;3(5):410-6.
245. Jin R, Zhou W. TIF-IA: An oncogenic target of pre-ribosomal RNA synthesis. Biochim Biophys Acta. 2016;1866(2):189-96.
246. Bruna A, Nicolàs M, Muñoz A, Kyriakis JM, Caelles C. Glucocorticoid receptor-JNK interaction mediates inhibition of the JNK pathway by glucocorticoids. EMBO J. 2003;22(22):6035-44.
247. Mayer C, Bierhoff H, Grummt I. The nucleolus as a stress sensor: JNK2 inactivates the transcription factor TIF-IA and down-regulates rRNA synthesis. Genes & development. 2005;19(8):933-41.
248. Yoon M-S. mTOR as a Key Regulator in Maintaining Skeletal Muscle Mass. Frontiers in Physiology. 2017;8(788).
249. Mayer C, Zhao J, Yuan X, Grummt I. mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. Genes & development. 2004;18(4):423-34.

250. Iadevaia V, Zhang Z, Jan E, Proud CG. mTOR signaling regulates the processing of pre-rRNA in human cells. Nucleic Acids Res. 2012;40(6):2527-39.
251. Chen H. Population genetic studies in the genomic sequencing era. Dongwuxue Yanjiu. 2015;36(4):223-32.
252. Griffiths AJF MJ, Suzuki DT, et al. Population Genetics. An Introduction to Genetic Analysis 7th ed. New York: W. H. Freeman; 2000.
253. Platt A, Novembre J. A new era of human population genetics. Genome Biol. 2012;13(12):182-.
254. Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nature Communications. 2015;6(1):8018.
255. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-d7.
256. Takeuchi F, Katsuya T, Kimura R, Nabika T, Isomura M, Ohkubo T, et al. The fine-scale genetic structure and evolution of the Japanese population. PloS one. 2017;12(11):e0185487-e.
257. Klouwer FCC, Berendse K, Ferdinandusse S, Wanders RJA, Engelen M, Poll-The BT. Zellweger spectrum disorders: clinical overview and management approach. Orphanet J Rare Dis. 2015;10:151-.
258. Conzen SD. Recent advances in understanding glucocorticoid receptor function in cancer. Clin Adv Hematol Oncol. 2017;15(5):338-40.
259. Reyes-Aldasoro CC. The proportion of cancer-related entries in PubMed has increased considerably; is cancer truly "The Emperor of All Maladies"? PLoS One. 2017;12(3):e0173671-e.
260. Rose AJ, Vegiopoulos A, Herzig S. Role of glucocorticoids and the glucocorticoid receptor in metabolism: insights from genetic manipulations. The Journal of steroid biochemistry and molecular biology. 2010;122(1-3):10-20.
261. Dvorak Z, Pavek P. Regulation of drug-metabolizing cytochrome P450 enzymes by glucocorticoids. Drug metabolism reviews. 2010;42(4):621-35.
262. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. Mol Cell Endocrinol. 2011;335(1):2-13.
263. Henrie A, Hemphill SE, Ruiz-Schultz N, Cushman B, DiStefano MT, Azzariti D, et al. ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. Hum Mutat. 2018;39(8):1051-60.
264. Robinson K, Tiriveedhi V. Perplexing Role of P-Glycoprotein in Tumor Microenvironment. Front Oncol. 2020;10:265-.
265. Jacobson L. Hypothalamic-pituitary-adrenocortical axis: neuropsychiatric aspects. Comprehensive Physiology. 2014;4(2):715-38.
266. Hübner S, Dejager L, Libert C, Tuckermann JP. The glucocorticoid receptor in inflammatory processes: transrepression is not enough. Biological chemistry. 2015;396(11):1223-31.
267. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021;590(7845):290-9.
268. Devuyst O. The 1000 Genomes Project: Welcome to a New World. Perit Dial Int. 2015;35(7):676-7.
269. Ambrocio-Ortiz E, Pérez-Rubio G, Ramírez-Venegas A, Hernández-Zenteno R, Del Angel-Pablo AD, Pérez-Rodríguez ME, et al. Effect of SNPs in HSP Family Genes, Variation in the mRNA and Intracellular Hsp Levels in COPD Secondary to Tobacco Smoking and Biomass-Burning Smoke. Frontiers in Genetics. 2020;10(1307).

270. Leem AY, Park B, Kim YS, Jung JY, Won S. Incidence and risk of chronic obstructive pulmonary disease in a Korean community-based cohort. Int J Chron Obstruct Pulmon Dis. 2018;13:509-17.
271. Funatogawa I, Funatogawa T, Yano E. Trends in smoking and lung cancer mortality in Japan, by birth cohort, 1949-2010. Bull World Health Organ. 2013;91(5):332-40.
272. Cappello F, Macario AJL, Di Stefano A. Hsp27 and Hsp70 in chronic obstructive pulmonary disease: certainties vs doubts. Cell Stress Chaperones. 2015;20(5):721-3.
273. Holownia A, Mroz RM, Kielek A, Chyczewska E, Braszko JJ. Nuclear HSP90 and HSP70 in COPD patients treated with formoterol or formoterol and corticosteroids. Eur J Med Res. 2009;14 Suppl 4(Suppl 4):104-7.
274. Rao S, Yao Y, Ryan J, Li T, Wang D, Zheng C, et al. Common variants in FKBP5 gene and major depressive disorder (MDD) susceptibility: a comprehensive meta-analysis. Scientific reports. 2016;6:32687-.

Glossary and Acronyms

#

17OHP: 17α-hydroxyprogesterone|

A

ACTH: Adrenocorticotropic Hormone|
AD: Alzheimer's Disease| AF-1: Activation Function 1| AF-2: Activation Function 2| AI: Artificial Intelligence| AR: Androgen Receptor|
ARDS: Acute Respiratory Distress Syndrome| AVP: Arginine Vasopressin|

B

BAG-1: BAG Family Molecular Chaperone Regulator 1|

C

cAMP: cyclic AMP| CAR: Constitutive Androstane Receptor|
CDD: Conserved Domains Database| CEBPB: CCAAT/Enhancer‐Binding Protein Beta| CEs: catecholamines| CNS: Central Nervous System| CoFs: Cofactors|
COPD: Chronic Obstructive Pulmonary Disease| CORT: Cortisol|
COUP-TFa: Chicken Ovalbumin Upstream Promoter-Transcription Factor alpha| COUP-TFb: Chicken Ovalbumin Upstream Promoter-Transcription Factor beta| CREs: Cis-Regulatory Elements| CRH: Corticotropin-Releasing Hormone|
CTD: C-Terminal Domain| CYPs: Cytochrome P450 enzymes|

D

DAX1: Dosage Sensitive Sex Reversal, Adrenal Hypoplasia Congenita Critical Region on the X Chromosome, Gene 1|
DBD: DNA-Binding Domain| DNMTs: DNA methyltransferases|

E

E: Epinephrine| EAR2: V-erbA-related Protein 2| EHRs: Electronic Health Records| ERa: Estrogen Receptor alpha| ERb: Estrogen Receptor beta|
ERRa: Estrogen-Related Receptor alpha| ERRb: Estrogen-Related Receptor beta| ERRg: Estrogen-Related Receptor gamma|

F

FKBPs: FK506-binding Proteins|FXR: Farnesoid X Receptor| FXRb: Farnesoid X receptor beta|

G

GAA: Acid Alfa-Glucosidase| GCNF: Germ Cell Nuclear Factor| GCs: Glucocorticoids| GH: Growth Hormone| GnRH: Gonadotropin-Releasing Hormone|
GR: Glucocorticoid Receptor| GRE: Glucocorticoid Response Element|
GTFs: General Transcription Factors| GWAS: Genome-Wide Association Studies|

H

HD: Homeodomain| HMG: High Mobility Group Box|
HNF4a: Hepatocyte Nuclear Factor 4 alpha|
HNF4g: Hepatocyte Nuclear Factor 4 gamma| Hop: Hsp70-Hsp90 Organizing Protein|
HPA axis: Hypothalamic-Pituitary-Adrenal axis|
HPG axis: Hypothalamic-Pituitary-Gonadal Axis| HR: Hinge Region|
HSD3B2: 3-beta-hydroxysteroid Dehydrogenase| Hsp40: Heat Schock Protein 40 kDa| Hsp70: Heat Shock Protein 70kDa| Hsp90: Heat Shock Protein 90 kDa|

I

IL-6: Interleukin-6|

J

JNK: c-Jun N-terminal kinases|

L

LBD: Ligand-Binding Domain| LC: Locus Coeruleus|
LRH-1: Liver Receptor Homolog-1| LXRa: Liver X Receptor alpha|
LXRb: Liver X Receptor beta|

M

MDD: Major Depressive Disorder| MDM2: Murine Double Minute 2| miRNAs: MicroRNAs| miRs: MicroRNAs| MMP2: Matrix Metalloproteinase 2|
MMP9: Matrix Metalloproteinase 9| MOE: Molecular Operating Environment| MOF: Mometasone Furoate| MR: Mineralocorticoid Receptor|
MS: Multiple Sclerosis| mTOR: mammalian target of rapamycin|

N

ncRNA: Non-Coding RNA| NE: Norepinephrine| NF-kB: Nuclear Factor Kappa Beta|
NGF1B: Nerve Growth Factor 1B| nGRE: Negative Glucocorticoid Response Element|
NGS: Next-Generation Sequencing| NOR1: Neuron-Derived Orphan Receptor 1|
NPC: Nuclear Pore Complex| NRs: Nuclear Receptors| NTD: N-Terminal Domain|
NURR1: Nurr-Related Factor 1|

P

PCOS: Polycystic Ovary Syndrome| PD: Parkinson's Disease|
PDB: RSCB Protein Data Bank| PFC: Prefrontal Cortex| PIC: Pre-Initiation Complex|
PKA: Protein Kinase A| PNR: Photoreceptor-Cell-Specific Nuclear Receptor| POMC: Pro-Opiomelanocortin|
PPARa: Peroxisome Proliferator-Activated Receptor alpha| PPARb: Peroxisome Proliferator-Activated Receptor beta|
PPARg: Peroxisome Proliferator-Activated Receptor gamma|
PR: Progesterone Receptor| PRRs: Pattern Recognition Receptors|
PTGES3/p23: Prostaglandin E Synthase 3| PTMs: Post-Translational Modifications|
PTSD: Post-Traumatic Stress Disorder| PVN: Paraventricular Nucleus|
PXR: Pregnane X Receptor|

R

RA: Rheumatoid Arthritis| RARa: Retinoic Acid Receptor alpha|
RARb: Retinoic Acid Receptor beta| RARg: Retinoic Acid Receptor gamma| REs: Response Elements| Rev-Erba: Rev-Erb alpha| Rev-Erbb: Rev-Erb beta|
RhoA: Ras Homolog Family Member A| RMSD: Root Mean Square Deviation| RNAIIP: RNA-Polymerase 2| RORa: RAR-related Orphan Receptor alpha| RORb: RAR-related Orphan Receptor beta|
RORg: RAR-related Orphan Receptor gamma| RXRa: Retinoid X Receptor alpha| RXRb:Retinoid X Receptor beta| RXRg: Retinoid X Receptor gamma|

S

SAM system: Sympathetic Adreno-Medullary system| SF1: Steroidogenic Factor 1|
SGK1: Serum and Glucocorticoid-regulated Kinase 1|
SHP: Short Heterodimeric Partner| SNPs: Single Nucleotide Polymorphisms|

T

T2DM: Typed 2 Diabetes Mellitus| TFs: Transcription Factors|

THRa: Thyroid Hormone Receptor alpha| THRb: Thyroid Hormone Receptor beta| TLX: Tailless Homolog Orphan Receptor| TR2: Testicular Receptor 2| TR4: Testicular Receptor 4| TSH: Thyroid-Stimulating Hormone|

U

UPGMA: Unweighted Pair-Group Method| USP: Ultraspiracle Protein|

V

VDR: Vitamin D Receptor| VEGF: Vascular Endothelial Growth Factor|

List of Publications

Salis C, Papakonstantinou E, Pierouli K, **Mitsis A (as Mitsis Athanasios)**, Basdeki L, Megalooikonomou V, Vlachakis D, Hagidimitriou M. A genomic data mining pipeline for 15 species of the genus Olea. EMBNet J. 2019, e922

Mitsis T, Papageorgiou L, Efthimiadou A, Bacopoulou F, Vlachakis D, Chrousos GP, Eliopoulos E. Transcription Factors and Evolution: An integral part of gene expression. World Academy of Sciences Journal. 2020, 2(1), 3-8 https://doi.org/10.3892/wasj.2020.32

Mitsis T, Papageorgiou L, Efthimiadou A, Bacopoulou F, Vlachakis D, Chrousos GP, Eliopoulos E. A comprehensive structural and functional analysis of the Ligand Binding Domain of the Nuclear Receptor Superfamily shows highly conserved signaling motifs and two distinct canonical forms through evolution. World Academy of Sciences Journal. 2020; 1(6), 264-274 https://doi.org/10.3892/wasj.2020.30

Vlachakis D, Papakonstantinou E, **Mitsis T**, Pierouli K, Diakou I, Chrousos GP, Bacopoulou F. Molecular mechanisms of the novel coronavirus SARS-CoV-2 and potential anti-COVID19 pharmacological targets since the outbreak of the pandemic Food Chem Toxicol 2020, 2020 Dec;146:111805. https://doi.org/10.1016/j.fct.2020.111805

Mitsis T, Pierouli K, Diakou KL, Papakonstantinou E, Bacopoulou F, Chrousos GP, Vlachakis D. Exosomics EMBnet J 2020;26:e934

Vlachakis D, **Mitsis T**, Nicolaides N, Efthimiadou A, Giannakakis A, Bacopoulou F, Chrousos GP Exosomal Endocrinology: Functions, Pathophysiology, and Current Insights Molecular Medicine Reports 2021;23(1):26. doi: 10.3892/mmr.2020.11664.

Papakonstantinou E, Diakou KI, **Mitsis T**, Dragoumani K, Bacopoulou F, Megalooikonomou V, Kossida S, Chrousos GP, Vlachakis D Molecular fusion events in carcinogenic organisms: a bioinformatics study for the detection of fused proteins between viruses, bacteria and eukaryotes EMBnet Journal 2021, Accepted. In Press
Papakonstantinou E, **Mitsis T**, Dragoumani K, Bacopoulou F, Megalooikonomou V, Chrousos GP, Vlachakis D The medical cyborg concept EMBnet Journal 2021, Accepted. In Press

Diakou KI, **Mitsis T**, Pierouli K, Papakonstantinou E, Bongcam-Rudloff E, Wayengera M, Vlachakis D. Ebola virus disease and current therapeutic strategies; a review. Adv Exp Med Biol 2021, Accepted. In Press

Diakou KI, **Mitsis T**, Pierouli K, Papakonstantinou E, Megalooikonomou V, Efthimiadou A, Vlachakis D Study of the Langat virus RNA-dependent RNA polymerase through homology modeling. EMBnet Journal 2021;26:e944

Papakonstantinou E, Pierouli K, **Mitsis T**, Diakou KI, Palaiogeorgou AM, Bacopoulou F, Chrousos GP, Eliopoulos E, Vlachakis D COVID‐19 global social lockdowns: Energy‐related, psychological, epigenetic, health and environmental impacts (Review) International Journal of Epigenetics 2021, 1, 6. https://doi.org/10.3892/ije.2021.6

Papakonstantinou E, **Mitsis T**, Dragoumani K, Efthimiadou A, Bacopoulou F, Chrousos GP, Eliopoulos E, Vlachakis D Materials of biological origin and biofuels: Small environmental footprint and epigenetic impact (Review) International Journal of Epigenetics 2021, 1, 8. https://doi.org/10.3892/ije.2021.8

Papakonstantinou E, Dragoumani K, Efthimiadou A, Palaiogeorgou AM, Pierouli K, **Mitsis T**, Chrousos GP, Bacopoulou F, Vlachakis D Haematological malignancies implications during the times of the COVID-19 pandemic (Review) Oncology Letters 2021, Accepted, In Press

Leptidis S, Papakonstantinou E, Diakou KI, Pierouli K, **Mitsis T**, Dragoumani K, Bacopoulou F, Sanoudou D, Chrousos GP, Vlachakis D Epitranscriptomics of Cardiovascular Diseases International Journal of Molecular Medicine 2021, Accepted, In Press