Spring 2018 – Systems Biology of Reproduction
Discussion Outline – Fetal Development & Birth Systems
Michael K. Skinner – Biol 475/575
CUE 418, 10:35-11:50 am, Tuesday & Thursday
April 19, 2018
Week 15

Fetal Development & Birth Systems

Primary Papers:

1. Renthal, et al. (2013) Nature Reviews Endocrinology 9:391-401
2. Glotov, et al. (2015) BMC Systems Biology 9(Suppl 2):S4
3. Feltes, et al. (2013) PLoS ONE 8(4)e61743

Discussion

Student 15: Reference 1 above

- What are the target organs for the miRNA in pregnancy?
- What gene clusters and networks are regulated?
- What progesterone targets and impacts influence contractility?

Student 16: Reference 2 above

- What major diseases are compared with preeclampsia and why?
- What networks were identified and impact?
- What risk factors were identified?

Student 18: Reference 3 above

- What impact does smoking have on embryonic development?
- What technical approach was used and types of networks investigated?
- What signaling pathways and cellular processes are effected and associated networks?

MicroRNAs—mediators of myometrial contractility during pregnancy and labour

Nora E. Renthal, Koriand'r C. Williams and Carole R. Mendelson

Abstract | The maintenance of myometrial quiescence and initiation of contractility, which lead to parturition at term and preterm, involve a shifting equilibrium between anti-inflammatory and proinflammatory signalling pathways. Progesterone (P_4), acting through the progesterone receptor (PR), has an essential and multifaceted role in the maintenance of myometrial quiescence. This effect of P_4–PR signalling is mediated, in part, by its anti-inflammatory actions and capacity to repress the expression of genes that encode proinflammatory cytokines, such as IL-1 and IL-6, and contraction-associated proteins, such as *OXTR*, *GJA1* and *PTGS2*. By contrast, increased expression of genes that ultimately lead to parturition is mediated by enhanced inflammatory and estradiol-17β (E_2) and estrogen receptor α signalling, which reduce PR function, thus further intensifying the inflammatory response. To obtain a more complete understanding of the molecular events that underlie the transition of the pregnant myometrium from a refractory to a contractile state, the roles of microRNAs, their targets, and their transcriptional and hormonal regulation have been investigated. This article reviews the actions of the miR-200 family and their P_4-regulated targets—the transcription factors ZEB1, ZEB2 and STAT5B—in the pregnant myometrium, as well as the role of miR-199a-3p and miR-214 and their mutual target PTGS2. The central role of ZEB1 as the mediator of the opposing actions of P_4 and E_2 on myometrial contractility will be highlighted.

Renthal, N. E. *et al. Nat. Rev. Endocrinol.* **9**, 391–401 (2013); published online 14 May 2013; doi:10.1038/nrendo.2013.96

Introduction

Each year, 15 million babies are born prematurely throughout the world (~11% of all live births); >1 million of these babies die because of complications of their prematurity. Preterm birth, defined as birth at <37 weeks of gestation, is the leading cause of neonatal death and the second leading cause of death in children under the age of 5 years worldwide.[1] The rate of preterm birth ranges from 5% in several European countries to 18% in parts of sub-Saharan Africa. In the USA, the incidence of preterm birth has steadily increased over the past two decades and has to date levelled off at ~12% of all live births. Notably, the incidence of preterm birth in the USA is higher among certain racial and ethnic groups compared with the general population. For example, the prematurity rate in black neonates is ~18%, whereas ~11% of white neonates are born prematurely. The reasons for this racial disparity are not understood.[2] To elucidate the underlying pathogenesis and prevent this high incidence of preterm birth, the signalling pathways that maintain quiescence of the myometrium throughout pregnancy and mediate its conversion into a synchronously contractile unit must be understood.

The maintenance of myometrial quiescence and the timing of labour involve a delicate balance between hormonal, inflammatory and physical factors that regulate integrated signalling pathways between the mother and the fetus.[3,4] MicroRNAs (miRNAs) and their regulation serve a pivotal role in the molecular events that underlie the transition of the pregnant myometrium from a refractory to a contractile state during term and preterm labour. These evolutionarily conserved regulators of gene expression play important parts in a variety of biological and pathological processes, including cell differentiation,[5–7] cancer,[8,9] immune regulation[10] and female reproduction.[11–31]

Consequently, miRNAs might serve as hormonally modulated mediators of inflammation-associated and contraction-associated gene expression in the pregnant uterus from mice to humans. Here, we review the current knowledge on the hormonal regulation of myometrial quiescence and contractility, with a focus on the role of miRNAs and their hormonal regulation during pregnancy and labour.

Myometrial quiescence during pregnancy

Quiescence of the myometrium throughout most of pregnancy is maintained by increased circulating levels of progesterone (P_4), which acts via the nuclear progesterone receptor (PR).[3] The PR maintains uterine quiescence, in part, by tethering to and antagonizing the actions of the transcription factors NF-κB and AP-1 on the expression of genes that encode proinflammatory cytokines, such as IL-1, IL-6 and IL-8, and chemokines, such as CCL2.[32–35] P_4–PR signalling also blocks NF-κB activation by increasing expression of the NF-κB inhibitor IκBα.[32]

Competing interests
The authors declare no competing interests.

Department of Pediatrics, Children's Medical Center Dallas, 1935 Medical District Drive, Dallas, TX 75235, USA (**N. E. Renthal**). Departments of Biochemistry and Obstetrics and Gynecology, the University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA (**K. C. Williams, C. R. Mendelson**).

Correspondence to: C. R. Mendelson carole.mendelson@utsouthwestern.edu

Key points

- Progesterone (P_4), acting via the progesterone receptor (PR), maintains uterine quiescence in part by increasing the expression of *ZEB1* and *ZEB2*, which inhibit the contraction-associated genes *OXTR* and *GJA1*
- The initiation of myometrial contractility is mediated by an increased inflammatory response, an associated increase in 17β-estradiol (E_2) and estrogen receptor (ERα) signalling and a decline in PR function
- Near term, the myometrial inflammatory response is promoted by physical and hormonal signals from mother and fetus; in preterm labour, the increased inflammatory response is commonly induced by a bacterial infection
- Expression of the miR-200 family increases in mouse and human myometrium near term and suppresses ZEB1 and ZEB2 levels, which results in the de-repression of contractile genes and increased myometrial contractility
- Increased miR-200 expression near term also inhibits STAT5B; decreased STAT5B levels de-repress 20α-hydroxysteroid dehydrogenase and increase myometrial metabolism of P_4
- Increased E_2–ERα signalling and the decline in PR function near term mediate decreased expression of ZEB1, and of miR-199a-3p and miR-214, which contributes to the induction of PTGS2

Furthermore, the PR prevents myometrial contractility by increasing expression of the zinc finger E-box-binding homeobox 1 transcription factor ZEB1, which inhibits expression of genes that encode contraction-associated proteins, such as oxytocin receptor, gap junction α-1 protein (GJA1, also known as connexin 43), and prostaglandin G,H synthase 2 (PTGS2, also known as cyclooxygenase 2), the critical enzyme in the synthesis of contractile prostaglandins, such as prostaglandin $F_2\alpha$.[29,31] Thus, PR maintains myometrial quiescence mainly by inhibiting the expression of genes associated with inflammation and contraction (Figure 1).

Initiation of labour

The inflammatory response

The initiation of labour, both at term and preterm, is associated with an upregulated inflammatory response, which is characterized by increased levels of proinflammatory cytokines in the amniotic fluid[36] and infiltration of the myometrium, cervix, and fetal membranes by neutrophils and macrophages.[37–39] In preterm labour, the intra-amniotic bacterial infection that is associated with chorioamnionitis (inflammation of the fetal membranes) can lead to increased cytokine levels in amniotic fluid and induce inflammatory cell migration.[40] By contrast, near term, increased mechanical stretch[35,41] and hormonal factors produced by the developing fetus[38,42–45] provide the inflammatory stimuli. The invading immune cells secrete cytokines and chemokines[46] that promote the activation of NF-κB and other inflammation-associated transcription factors, such as AP-1, in the myometrium[38,47] and cervix.[48–50] These activated transcription factors, in turn, inhibit PR function, which further induces the inflammatory response and expression of myometrial contractile genes,[51–54] culminating in parturition (Figure 2).

Parturition—a decline in PR function

In rodents, maternal P_4 levels in the circulation decrease precipitously near term.[55] This finding led to the hypothesis that labour at term is associated with P_4 withdrawal. By contrast, in humans and in guinea pigs, maternal circulating P_4 levels, as well as myometrial PR levels, remain elevated throughout pregnancy and into labour.[44] Nevertheless, treatment with PR antagonists, such as mifepristone (also referred to as RU-486) or onapristone, can increase cervical ripening and induce spontaneous labour or increase the sensitivity to labour induction by oxytocin or prostaglandins.[56–59] Of note, even in mice, maternal P_4 levels at term remain well above the dissociation constant (K_d) for binding to PR.[60] Thus, the progression to labour at term, which is associated with an increased inflammatory response to mechanical and

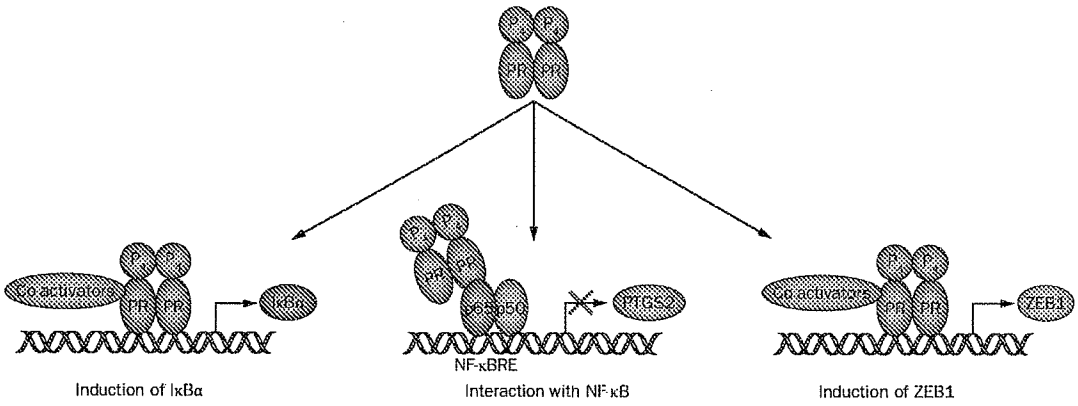

Figure 1 | P_4–PR regulation of myometrial quiescence. PR maintains myometrial quiescence, in part, by blocking activation of NF-κB and preventing its transcriptional activation of proinflammatory genes, such as *PTGS2*. The P_4–PR complex exerts this action, in part, by increasing expression of the NF-κB inhibitor IκBα, which prevents activation and nuclear translocation of the NF-κB subunits p50 and p65. PR also inhibits activation of contraction-associated genes such as *PTGS2* by interaction with NF-κB, that is, the P_4–PR complex tethers to NF-κB bound to response elements in the promoter of *PTGS2*. In addition, PR prevents myometrial contractility by increasing protein expression of ZEB1, which inhibits expression of contraction-associated genes, for example, *OXTR*, *GJA1* and *PTGS2*, by interaction with their promoter regions. Abbreviations: IκBα, NF-κB inhibitor α; NF-κB, nuclear factor κB; NF-κBRE, NF-κB response element; P_4, progesterone; PR, progesterone receptor; PTGS2, prostaglandin G,H synthase 2; ZEB1, zinc finger E-box-binding homeobox 1.

hormonal signals from the mother[61] and fetus,[38,42,62,63] might be mediated by a chain of molecular events that impair the ability of P_4 via PR to regulate target genes in the uterus that maintain myometrial quiescence.

The decline in myometrial PR function near term is caused by numerous factors, which include: a decrease in PR co-activators, for example, CREB-binding protein and the steroid receptor co-activators SRC-2 and SRC-3;[64] increased expression of inhibitory and truncated PR isoforms;[47,65] antagonism of the PR by NF-κB,[32,66,67] which is activated in the myometrium near term;[38,47] and enhanced local metabolism of P_4 to inactive products. The breakdown of P_4 into inactive metabolites is mediated by 20α-hydroxysteroid dehydrogenase (20α-HSD), an enzyme encoded by the *AKR1C1* gene in humans and the *Akr1c18* gene in mice, in the uterus[30] and by steroid 5α-reductase type 1, which is encoded by the *SRD5A1* gene, in the cervix (Figure 2).[68–70] Mice with a targeted deletion in *Srd5a1* exhibit defects in cervical ripening, which prevent delivery.[68,69] The inhibition of delivery occurs despite a precipitous decline in the levels of maternal circulating P_4 near term, which suggests that, in the mouse, local metabolism of P_4 in the cervix is also required for P_4 withdrawal and the initiation of parturition. Mice that lack 20α-HSD also manifest markedly delayed parturition,[71,72] despite a marked decline in circulating P_4 levels.[71]

Importantly, increased circulating estradiol-17β (E_2) levels[73,74] and enhanced estrogen receptor α (ERα) activity[65,75–77] near term also promote a cascade of pro-inflammatory events that contribute to the decline in PR function and initiate labour. Estrogens induce an influx of macrophages and neutrophils into the uterus and antagonize the anti-inflammatory actions of P_4 acting through PR.[65,78] Furthermore, ERα activation facilitates labour by enhancing transcription of the contraction-associated genes *OXTR*,[79] *GJA1*[80] and *PTGS2*,[31,65] which are expressed at low to undetectable levels in the uterus throughout most of pregnancy but are highly upregulated at term.[31,81,82]

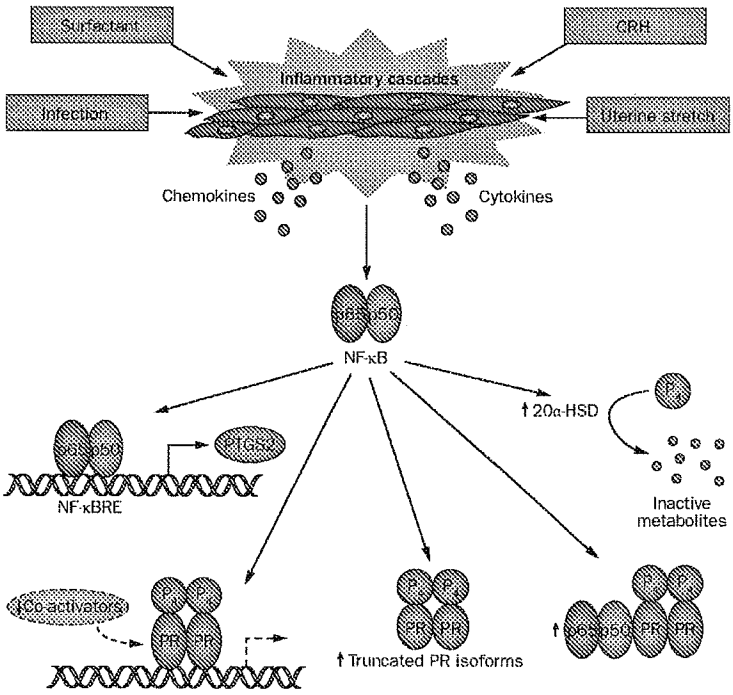

Figure 2 | The increased inflammatory response associated with parturition. Near term, a number of fetal and maternal signals, including secretion of surfactant by the fetal lung, production of hormonal factors by the placenta (such as corticotropin-releasing hormone), and mechanical stretch of the uterus, result in increased levels of proinflammatory cytokines in amniotic fluid and infiltration of the myometrium, cervix and fetal membranes by neutrophils and macrophages. The invading immune cells secrete cytokines and chemokines, promoting activation and binding of the NF-κB subunits p65 and p50 to the promoter of *PTGS2*. The increased inflammatory response can also inhibit PR function by suppressing expression of co-activators (such as CREB-binding protein and the steroid co-activators SRC-2, SRC-3), upregulating expression of truncated PR isoforms, direct interaction of NF-κB with PR, and induction of enzymes that metabolize P_4 to inactive products, such as 20α-HSD and steroid 5α-reductase type 1. Collectively, these processes result in a further induction of the inflammatory response and increased expression of myometrial contractile genes, culminating in parturition. Abbreviations: 20α-HSD, 20α-hydroxysteroid dehydrogenase; NF-κB, nuclear factor κB; NF-κBRE, NF-κB response element; P_4, progesterone; PR, progesterone receptor; PTGS2, prostaglandin G,H synthase 2.

MicroRNAs in female reproductive biology

Since their discovery almost two decades ago, miRNAs have increasingly been recognized as potent, evolutionarily conserved, post-transcriptional regulators of gene expression (Box 1). miRNAs exhibit important regulatory roles in vascular smooth muscle[83] and in female reproductive tissues, where they have been implicated in proliferation, differentiation, embryo implantation, hormone responsiveness, parturition and pathology (Table 1).[12–31,84–89] For example, conditional deletion of *Dicer1* in ovarian granulosa cells and in derivatives of the Müllerian duct (that is, oviduct, uterus and cervix) caused malformations of the uterus and sterility in mice.[14,22,24,27] Histologic examination revealed a decreased myometrial layer and fewer uterine glands. Although these findings are indicative of the importance of miRNAs for global uterine form and function, they shed little light on the roles of individual miRNAs in specific reproductive functions.

Coordinated miRNA and gene expression microarray analyses were used to identify miRNAs and targets that are differentially regulated in quiescent compared with contractile pregnant mouse myometrium.[29] Myometrial tissues from pregnant mice were removed at 15.5 days *post coitum* (dpc), when the myometrium is quiescent, and at 18.5 dpc, just before labour, which typically begins at 19 dpc. In total, 15 miRNAs were found to be significantly upregulated and downregulated ($P<0.05$), in association with a body of gestationally regulated genes.[29] Among the most markedly upregulated miRNAs was a conserved family, the miR-200 family, which was highly induced at term.[29] miRNAs that were significantly downregulated between 15.5 dpc and 18.5 dpc included the *miR-199a/214* cluster.[29,31]

The miR-200 family and its targets

The miR-200 family is comprised of five miRNAs arranged into two conserved clusters (*miR-200b/200a/429*

Box 1 | MicroRNAs—evolutionarily conserved, potent regulators of gene expression

The majority of mammalian microRNAs (miRNAs) exist throughout the genome as promoter-driven RNA polymerase II transcribed genes. Their biogenesis begins in the nucleus where they are transcribed as part of long RNA transcripts (pri-miRs). miRNAs commonly exist in polycistronic clusters, wherein multiple miRNA genes are encoded in a single primary transcript. These clustered miRNAs are often structurally similar to one another and have related biological functions. The pri-miRs are processed in the nucleus to ~70-nucleotide hairpin structures (pre-miRs) by Drosha/DGCR8, an RNase III endonuclease, and then exported to the cytoplasm where they are further processed by the RNase III Dicer to form a short miRNA duplex. Upon loading of one of the miRNA strands onto the RNA-induced silencing complex (RISC), containing Argonaute (Ago) proteins, the miRNA directs the RISC to its mRNA targets. Typically, only one of the two strands is selected as the mature, RISC-associated miRNA, whereas the other is either degraded, or expressed to a lesser extent. In some cases, both strands of the miRNA duplex become fully mature miRNAs and are incorporated into RISC complexes. In this case, the miRNAs are termed miR-XXX-3p and miR-XXX-5p. Approximately one third of mammalian miRNAs are encoded within introns of protein-coding 'host' genes and are generated upon processing of the host RNA transcripts. The intron-encoded miRNAs are under control of the 'host' gene promoter and commonly mediate a related biological function to the host protein. This process has been reviewed in detail elsewhere.[126]

miRNA expression can be regulated at the transcriptional level[3] and at the level of their processing.[127] miRNAs typically function to repress gene expression by binding through imperfect base pairing via their seed sequences (nucleotide 2–8 at the 5′-end) to complementary sites (seed match), which typically exist within the 3′ untranslated region of target mRNAs. Binding of the miRNA results in degradation of the mRNA target and/or inhibition of translation.[128,129] miRNAs can act as rheostats or as on–off switches of gene expression. The effect of a specific miRNA on the regulation of a target can be subtle, whereas the combined actions of several related miRNAs on the same target can have a pronounced phenotypic consequence. This phenomenon is especially true when miRNAs with similar or identical seed sequences are expressed as part of the same polycistronic transcript and/or when miRNAs encoded within the same transcript have different seed sequences that bind to the same target mRNA. It is estimated that ~1,000 miRNAs are encoded by the human genome and that these regulate approximately one-third of the expressed human genes.[5]

and *miR-141/200c*) that exist on two different chromosomes within the mouse and human genomes.[29] Each cluster is <2 kbp in length; the miRNAs in each cluster are coordinately transcribed.[90] miR-200b, miR-200c and miR-429 contain an identical seed sequence; miR-141 and miR-200a also share an identical seed sequence that differs by only a single nucleotide from that of miR-200b, miR-200c and miR-429 (Figure 3). Thus, all members of the miR-200 family probably share targets.

miR-200 targets: ZEB1 and ZEB2

From the results of a gene expression array and with the use of prediction algorithms[91] and published findings regarding miRNA–mRNA-target relationships, our group identified a pool of regulated miR-200 targets among the mRNAs downregulated at term in mice. The two most significantly downregulated miR-200 targets were the transcriptional repressors Zeb1 and Zeb2.[29] Importantly, the *Zeb1* gene was previously reported to be highly expressed in mouse myometrium, and its expression was upregulated by P_4 and PR signalling.[92,93] In myometrial tissues of pregnant mice, miR-200a,[30] miR-200b and miR-429[29] expression was significantly increased after 17.5 dpc; this timing corresponded with a reduction in Zeb1 and Zeb2 levels in the myometrium.[29] Importantly, the increased expression of miR-200b and miR-429 and decreased levels of ZEB1 and ZEB2 were also observed in the myometrium of women in labour compared with pregnant women not in labour.[29] This finding indicates that the relationship between miR-200 family members and ZEB1 and ZEB2 is conserved between mice and humans.

ZEB1 and ZEB2 were previously shown to be repressed by members of the miR-200 family in a double-negative feedback loop.[90,94–96] Furthermore, ZEB1 and ZEB2 were shown to stimulate epithelial to mesenchymal transition in cancer cells.[96–98] A clear inverse relationship exists between ZEB1 and ZEB2 and miR-200 family members in myometrium; overexpression of mimics of miR-200b and miR-429 in immortalized human myometrial cells reduced ZEB1 and ZEB2 levels, whereas transduction of mouse myometrial cells in primary culture with recombinant adenoviruses containing *Zeb1* and *Zeb2* expression vectors repressed expression of miR-200 and miR-429.[29] Moreover, using quantitative chromatin immunoprecipitation PCR (ChIP-qPCR), *in vivo* binding of Zeb1 to the *miR-200b/200a/429* promoter in mouse myometrium was observed at fairly high levels at 15.5 dpc; Zeb1 binding declined markedly at 18.5 dpc in association with the increase in miR-200b and miR-429 expression.[29]

miR-200b and miR-429 levels were increased and Zeb1 and Zeb2 mRNA and protein levels were decreased[29] in pregnant myometrium in two mouse models of premature parturition, induced either by a single subcutaneous injection of the antiprogestin and antiglucocorticosteroid mifepristone[99] or by intra-amniotic injection of the bacterial endotoxin, lipopolysaccharide. Conversely, daily injection of P_4 into pregnant mice from 15.5 dpc to 18.5 dpc, which inhibited myometrial contractile gene expression and delayed labour, specifically induced *Zeb1* expression. Surprisingly, P_4 injection had no effect on *Zeb2* expression.[29]

ZEB1 and *ZEB2* are expressed at relatively high levels in the myometrium throughout most of pregnancy, which raised the question regarding what factor(s) cause the pregnancy-associated induction of ZEB2. In studies using cultured mouse myometrial cells, *Zeb1* overexpression caused a time-dependent upregulation of Zeb2 levels.[29] This finding suggests that induction of *Zeb1* expression by P_4 and the associated inhibition of miR-200 family members, in turn, relieves suppression of *Zeb2*, allowing its subsequent induction. Notably, *Zeb1* has two putative P_4 response elements (PREs) in its promoter, whereas none are apparent in the *Zeb2* promoter. Accordingly, in co-transfection studies of human embryonic kidney (HEK293) cells with a *Zeb1*-luciferase reporter construct containing 978 bp of the *Zeb1* 5′-flanking sequence, which includes the two putative PREs, *Zeb1* promoter activity was induced by co-transfection of wild-type PR, but not by a mutant PR that contains a mutation in the DNA-binding domain.[29]

ZEB1 and ZEB2 suppress OXTR and GJA1

Two genes known to be upregulated near term in myometrium of a variety of species are *OXTR* and *GJA1*.[52,53,100–103] As *ZEB1* and *ZEB2* expression declined, *OXTR* and *GJA1* expression was temporally upregulated in mouse

and human myometrium.[29] GJA1 is a component of gap junctions in the myometrium, which mediate intercellular communication required for the synchronous contractions during labour. In mice, deletion of *Gja1* in smooth muscle leads to a substantial delay in the induction of labour.[104] Oxytocin (encoded by *OXT*) is widely accepted as a uterotonic agent; however, its role and that of the oxytocin receptor (encoded by *OXTR*) in parturition is uncertain; mice with a deletion in *Oxt* undergo normal parturition and give birth to live offspring,[105] and mice with a deletion of *Oxtr* show normal timing and duration of parturition.[106] These unexpected phenotypes might be due to a functional redundancy of the oxytocin signalling system and/or due to compensatory upregulation of other uterotonic systems, such as Ptgs2-mediated prostaglandin synthesis.

On the basis of the reciprocal temporal relationship between ZEB1, ZEB2 and contractile gene expression in pregnant mouse and human myometrium, we postulated that ZEBs suppress myometrial contractility by negatively regulating *OXTR* and *GJA1* expression. Indeed, overexpression of *ZEB1* or *ZEB2* in immortalized human myometrial cells caused a pronounced inhibition of OXTR and GJA1 mRNA levels.[29] Moreover, endogenous Zeb1 was bound at fairly high levels to E-box-containing regions of the mouse *Gja1* and *Oxtr* promoters at 15.5 dpc in pregnant mouse myometrium, whereas binding was markedly reduced at term.[29] To assess the functional roles of ZEB1 and ZEB2 on myometrial contractility, immortalized human myometrial cells transduced with *ZEB1* or *ZEB2* expression vectors or with control vectors were embedded in 3D collagen gels; the effects of oxytocin on contraction of the gels were analysed. Whereas oxytocin significantly induced contraction of collagen gel matrices embedded with cells transduced with control vectors, this action was blocked in cells transduced with *ZEB1* and *ZEB2* expression vectors. This finding indicates an inhibitory effect of these transcription factors on myometrial contractility *in vitro*.[29] Collectively, these findings support a role for the miR-200 family and ZEB1 and ZEB2 in the regulation of myometrial contractility during pregnancy and labour in mice and humans (Figure 4).

Throughout most of gestation, elevated circulating P_4 levels induce myometrial *ZEB1* expression via binding of P_4–PR to the *ZEB1* promoter. Elevated ZEB1 suppresses expression of *GJA1* and *OXTR*,[29] as well as the *miR-200* gene clusters[90,96] by binding to response elements within their promoters. The suppression of miR-200 promotes further upregulation of ZEB1 and increases expression of ZEB2. Together, ZEB1 and ZEB2 inhibit *GJA1* and *OXTR* to maintain myometrial quiescence. Near term, signals from mother and fetus, described above, together with a decline in circulating P_4,[55] and/or PR function result in an increased inflammatory response within the myometrium and a further decline in PR function. Consequently ZEB1 mRNA and protein are significantly reduced. The decline in ZEB1 enables upregulation of miR-200 expression, resulting in a further suppression of ZEB1 and inhibition of ZEB2. The combined decline in ZEB1 and ZEB2 enables marked upregulation of *OXTR* and *GJA1* gene expression and the induction of myometrial contractility (Figure 4).

Table 1 | MicroRNAs in pregnancy and labour

Tissue	Process	Subject	References
Endometrium	Embryo receptivity and implantation	miR-101a, miR-199a-3p let-7b miR-320, let-7a miR-222 miR-21 miR-96, miR-375, miR-219-5p and PR miRNA expression and endometrial receptivity miRNA and mRNA analysis with delayed implantation	Chakraborty *et al.*[17] Fu *et al.*[33] Xia *et al.*[19,20] Qian *et al.*[21] Hu *et al.*[28] Liu *et al.*[18] Li *et al.*[16] Su *et al.*[17]
Endometrium	Endometriosis	miR-21, Dicer miR-29c miR-222, miR-17-5p miR-9, miR-34 SNPs in miRNA target sequence Differential expression of miRNAs and pathway analysis	Aghajanova & Giudice[19] Hawkins *et al.*[95] Ramon *et al.*[98] Burney *et al.*[99] Zhao *et al.*[97] Ohlsson-Teague *et al.*[88]
Uterus	Development	Dicer knockout	Hawkins *et al.*[34] Nagaraja *et al.*[22] Gonzalez & Behringer[23] Hong *et al.*[27]
Myometrium	Pregnancy versus labour	miR-200 and ZEB miR-200 and STAT5B miR-199a-3p, miR-214	Renthal *et al.*[29] Williams *et al.*[30] Williams *et al.*[31]
Cervix	Pregnancy versus labour	miR-223, miR-34b, miR-34c	Hassan *et al.*[35]
Fetal membranes	Pregnancy versus term labour	miR-338, miR-223	Montenegro *et al.*[25,26] Kim *et al.*[24]

Abbreviation: miRNA, microRNA; PR, progesterone receptor; SNP, single nucleotide polymorphism.

miR-200a and STAT5B

As mentioned above, increased metabolism of P_4 within the uterus and cervix near term may contribute to a decline in PR function that is crucial for the initiation of parturition in all mammals.[30,68,69] In fact, increased P_4 metabolism in the uterus near term has been observed in a number of species.[107–110] In myometrium of pregnant women at term, a pronounced decrease in the ratio of P_4 to 20α-dihydroprogesterone was found;[110] 20α-dihydroprogesterone is an inactive metabolite of P_4 generated by 20α-HSD, a member of the aldo-ketoreductase (AKR) superfamily.[111] 20α-HSD is encoded by *AKR1C1* in humans and by *Akr1c18* in mice. Targeted deletion of *Akr1c18* in mice caused a pronounced delay in the initiation of labour.[72] Importantly, the transcription factor signal transducer and activator of transcription 5B (STAT5B) is a P_4-responsive transcriptional repressor of the gene that encodes 20α-HSD in reproductive tissues.[72,112] Consequently, *Stat5b* deficiency in mice resulted in increased expression of ovarian 20α-HSD, decreased circulating P_4 levels and caused abortion during mid-gestation.[72] Notably, the abortion rate in *Stat5b*-deficient mice was partially corrected by combined *Akr1c18* deficiency.[72]

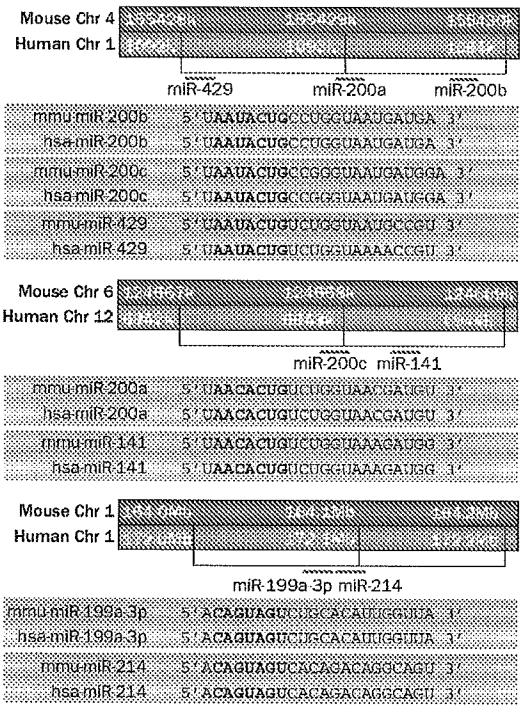

Figure 3 | Chromosomal location and seed sequences of members of the miR-200 family and the *miR-199a/214* cluster. The miR-200 family in mice and humans exists in two conserved clusters on mouse chromosome 4 and human chromosome 1 (*miR-200b/200a/429*) and on mouse chromosome 6 and human chromosome 12 (*miR-141/200c*), respectively. Family members are aligned according to their seed sequences (in bold), which are identical for miR-200b, miR-200c and miR-429 in mice and humans and differ by one nucleotide from miR-200a and miR-141, which are also identical and conserved in these species. Also shown are the locations of the *miR-199a/214* cluster in mouse and human chromosome 1 and the perfect conservation of the seed sequences of mouse and human miR-199a-3p and miR-214. Because these miRNAs are coordinately transcribed and contain distinct seed sequences, their interactions with different seed matches in the 3'-untranslated region of *PTGS2* can cooperatively exert a profound phenotypic effect. Abbreviations: Chr, chromosome; hsa, humans; miR, microRNA; mmu, mouse.

Expression of 20α-HSD in the ovarian corpus luteum remains low throughout pregnancy and increases in association with luteolysis at term;[113] hence, delayed labour in *Akr1c18*-deficient mice might have been caused by an inhibition of luteolysis and sustained circulating P_4 levels. However, in one of the *Akr1c18* knockout studies, labour was delayed in pregnant gene-targeted mice despite the fact that circulating P_4 levels declined in a similar manner as in wild-type mice.[71] This finding suggests that the decrease in ovarian P_4 production at term may not be sufficient and that actions of 20α-HSD to catalyse local metabolism of P_4 in the reproductive tract are also critical for the decline in PR function that leads to labour.

STAT5B is a target of miR-200a,[30] a member of the miR-200 family that increases dramatically at term in the myometrium of mice and humans. Remarkably, STAT5B mRNA and protein expression were reciprocally decreased in myometrial tissues of pregnant mice and humans, in association with the increase in miR-200a levels at term.[30] Moreover, these gestational changes were associated with decreased binding of endogenous Stat5b to response elements in the 5'-flanking region of the *Akr1c18* gene and an induction of 20α-HSD mRNA, protein and enzyme activity.[30] An intermediary role of miR-200a in P_4 induction of Stat5b was further suggested by the finding that P_4 treatment of ovariectomized mice inhibited miR-200a, increased Stat5b levels and inhibited 20α-HSD expression.[30] Conversely, injection of mifepristone increased miR-200a and 20α-HSD mRNA levels and inhibited *Stat5b* expression.[30] Given that the seed sequence of miR-200a is almost identical to that of miR-200b and miR-429, which are encoded within the same transcript,[29] these members of the miR-200 family might also target myometrial Stat5b and regulate 20α-HSD. However, this hypothesis remains to be confirmed. Moreover, miR-200a also targets ZEB1 and ZEB2 directly.[94] This finding suggests that increasing levels of miR-200a near term act cooperatively with miR-200b and miR-429 to inhibit ZEB expression and de-repress genes that encode proteins necessary for myometrial contractions.

These findings suggest that the elevated P_4 and PR activity in the myometrium during most of pregnancy causes induction of ZEB1 and inhibition of miR-200 expression, which is permissive for increased expression of *STAT5B*. Increased STAT5B levels maintain 20α-HSD at low concentrations and enable P_4 and PR function to remain elevated (Figure 4). The increased inflammatory response leading to term or preterm labour causes a decline in PR function and associated decrease in ZEB1 levels. This phenomenon permits upregulation of miR-200a and other miR-200 family members, an associated inhibition of *STAT5B* and induction of 20α-HSD expression and activity. The increased local metabolism of P_4 to inactive products in the myometrium contributes to the further decline in PR function and the progression to labour (Figure 4). These studies have revealed a robust positive feed-forward loop, wherein an initially modest decline in PR function and induction of miR-200 expression can escalate to an intensity that effectively reduces myometrial P_4 to levels below the K_d for binding to PR. This phenomenon culminates in a further decline of ZEB1 and ZEB2 levels, which permits the induction of contractile gene expression in response to increased NF-κB activation (Figure 4).

The *miR-199a/214* cluster and PTGS2

miRNAs that are significantly downregulated near term include miR-199a-3p, known to target the PTGS2 mRNA,[12] and miR-199a-5p, which inhibits NF-κB activation by targeting IKKβ.[114,115] Mature miR-199a-3p and miR-199a-5p are processed from the same precursor pri-miR-199a. Pri-miR-199a is synthesized as part of a 6-kb antisense transcript (Dnm3os) from the intron of the *Dnm3* gene.[116] The Dnm3os transcript, which is highly expressed in the uterus of pregnant mice,[116] also encodes miR-214, which targets PTEN[115] to activate the Akt pathway, as well as PTGS2.[31]

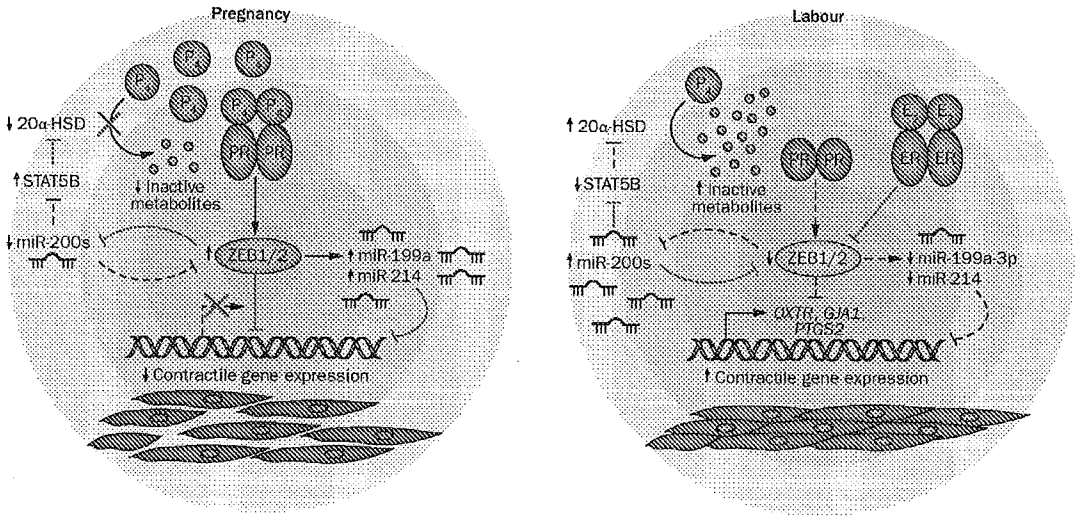

Figure 4 | Schematic representation of the pivotal role of ZEB1 and ZEB2 during pregnancy and labour. During pregnancy, increased levels of P_4 and increased PR function promote upregulation of ZEB1 levels in myometrium. ZEB1 inhibits expression of the miR-200 family and suppresses *OXTR* and *GJA1*. The decline in miR-200 levels further increases ZEB1 and ZEB2 levels, and these two proteins bind response elements upstream of the *miR-199a/214* cluster to enhance its expression, causing suppression of *PTGS2* expression and preventing synthesis of contractile prostaglandins. The decreased levels of miR-200s enable upregulation of another target, STAT5B, which inhibits expression of the gene encoding 20α-HSD, preventing the local metabolism of P_4 in myometrium. During the transition to labour, the decline in myometrial P_4 and PR function and increases in circulating E_2 and ERα activity cause downregulation of *ZEB1* expression. The decline in ZEB1 leads to induction of the miR-200 family, which further suppresses ZEB1 and ZEB2 levels, allowing upregulation of *OXTR* and *GJA1* expression. The decline in the levels of ZEB1 and ZEB2 also decreases expression of the *miR199a/214* cluster, which enables upregulation of *PTGS2* expression and increased synthesis of contractile prostaglandins. The increase in miR-200 expression inhibits STAT5B, permitting increased transcription of the gene that encodes 20α-HSD to promote increased metabolism of P_4 to inactive products in myometrium. Collectively, these molecular events contribute to the initiation of uterine contractility, leading to labour. Abbreviations: 20α-HSD, 20α-hydroxysteroid dehydrogenase; E_2, estradiol-17β; ER, estrogen receptor; GJA1, gap junction α-1 protein; miR, microRNA; OXTR, oxytocin receptor; P_4, progesterone; PR, progesterone receptor; PTGS2, prostaglandin G,H synthase 2; STAT5B, signal transducer and activator of transcription 5B; ZEB, zinc finger E-box-binding homeobox.

miR-214 was also observed to be expressed at high levels at 15.5 dpc and was downregulated in the uterus of pregnant mice at 18.5 dpc.[31] As discussed previously, NF-κB activation and expression of *PTGS2* increase in the pregnant myometrium during late gestation;[36,47,51] NF-κB signalling and PTGS-mediated prostaglandins are considered to be critical for the progression of labour. Whereas quantitative reverse transcription PCR of RNA isolated from myometrial tissues of pregnant mice at 15.5 dpc, 18.5 dpc and in active labour confirmed that miR-199a-3p and miR-214 were significantly downregulated in pregnant mouse myometrium at 18.5 dpc and during labour, a significant change in the expression of miR-199a-5p was not found.[31] For this reason, our studies have focused on the roles of miR-199a-3p and miR-214 in the regulation of PTGS2.

In pregnant mouse myometrium, Ptgs2 protein levels were significantly increased at 18.5 dpc and during labour, compared to 15.5 dpc, as miR-199a-3p and miR-214 levels reciprocally declined. By contrast, Ptgs2 mRNA levels remained low until 18.5 dpc and increased only during labour.[31] The relevance of the gestational increase in Ptgs2 protein levels is supported by the observation that prostaglandin $F_2\alpha$ levels were significantly increased in pregnant mouse uterus between 16 dpc and 18 dpc.[117] Our findings suggest that miR-199a-3p and miR-214 exert a direct effect on Ptgs2 mRNA translation rather than on mRNA stability.

This hypothesis was supported by the finding that overexpression of miR-199a-3p and miR-214 in cultured human myometrial cells decreased PTGS2 protein levels, but had no effect on PTGS2 mRNA levels.[31] Moreover, in myometrial samples from women in labour compared with those not in labour at term, in the absence of underlying infection, PTGS2 mRNA was unchanged, whereas PTGS2 protein levels were markedly increased during labour.[31] Thus, these collective findings suggest that PTGS2 expression in the pregnant myometrium is regulated at the level of mRNA translation and that the *miR-199a/214* cluster has an important role in this regulation. We view these findings to be relevant because the role of myometrial PTGS2 upregulation in the initiation of normal labour at term has been questioned.[118] This query was based on the observation that PTGS2 mRNA levels were not found to be increased in myometrial tissues from women in labour, compared with tissues from pregnant women not in labour, except in the presence of chorioamnionitis.[118] Notably, our studies revealed that myometrial levels of miR-199a-3p and miR-214 were significantly decreased in a mouse model of lipopolysaccharide-induced preterm labour, whereas Ptgs2

mRNA and protein levels were increased. Moreover, the physiological relevance of the relationship between the *miR-199a/214* cluster and PTGS2 in the regulation of myometrial contractility was further supported by the finding that miR-199a and miR-214 overexpression in cultured human myometrial cells blocked TNF-induced contractility to the same extent as the cyclooxygenase inhibitor indomethacin.[31]

As mentioned previously, P_4 and E_2 exert opposing effects on myometrial quiescence and contractility. Accordingly, E_2 treatment of ovariectomized mice suppressed, and P_4 treatment enhanced, uterine expression of miR-199a-3p and miR-214.[31] Interestingly, these opposing hormonal effects were found to be mediated by ZEB1, which is induced by P_4[29,92] and inhibited by E_2, and which activates transcription of miR199a and miR-214.[31] Thus, these findings have uncovered an intriguing pivotal role of ZEB1 as a negative regulator of the *miR-200b/200a/429* cluster and as a positive regulator of the *miR-199a/214* cluster that is under opposing control of P_4 and E_2 (Figure 4).

Future perspectives

The search for circulating biomarkers that are useful in predicting spontaneous preterm birth with the aim of clinical intervention has been ongoing for the past 40 years. In a review of the literature, of the 116 protein biomarkers analysed in 217 studies, not one provided predictive value for spontaneous preterm birth or yielded insight into the underlying pathophysiology.[119] The presence of specific miRNAs in blood and other body fluids suggests that they might serve as clinically useful biomarkers owing to a long half-life in body fluids, as well as the simplicity, accuracy and cost-effectiveness of their analysis.[120,121]

Moreover, chemically modified antisense oligonucleotides complementary to the seed sequence of miRNAs (anti-miRs or antagomiRs) have been used successfully *in vivo* to competitively inhibit miRNA function.[122–124] This finding suggests that miRNA manipulation has great potential as a future therapeutic strategy for the prevention of preterm birth. Given that members of an miRNA family share seed sequences, a single anti-miR can block the function of an entire family of miRNAs, making anti-miR therapy particularly appealing. Advances in the design of locked nucleic acid (LNA)-modified phosphorothioate oligonucleotides as highly stable miRNA antagonists that are taken up and retained by a variety of tissues for a period of up to 3 weeks[125] suggests a potential utility of these antagonists in the treatment or prevention of preterm labour and its potentially devastating consequences.

Conclusions

The studies described herein indicate that miRNAs have key collaborative roles in the hormonal control of myometrial quiescence and contractility during pregnancy and labour through regulation of contractile gene expression and PR function. They also suggest a central role for ZEB transcription factors and their unique capacity to act both as repressors and activators of gene transcription (Figure 4). Throughout most of pregnancy, when circulating P_4 levels are elevated and PR function is high, ZEB1 expression is induced to high levels. ZEB1 induction, in turn, causes suppression of the miR-200 family, which further increases ZEB1 levels and allows for de-repression of ZEB2. Together, ZEB1 and ZEB2 contribute to the maintenance of myometrial quiescence through their actions as potent transcriptional inhibitors of *OXTR* and *GJA1*. Elevated P_4–PR levels and suppression of the miR-200 family also enables upregulation of STAT5B levels, which inhibits expression of the P_4-metabolizing enzyme 20α-HSD. This inhibition permits myometrial tissue levels of P_4 to remain elevated, which increases PR function to sustain myometrial quiescence via the mechanisms just described, and suppresses NF-κB activation of inflammation-associated genes in the myometrium, such as *PTGS2*. Increased levels of ZEB1 and ZEB2 also cause upregulation of miR-199a-3p and miR-214, which directly target PTGS2 mRNA, resulting in maintained suppression of contractile prostaglandin synthesis.

Near term, the increased inflammatory response caused by uterine stretch, increased E_2 production and/or ERα activity and other signalling molecules produced by mother and fetus initiate a decline in PR function. This decrease further enhances NF-κB activation and the inflammatory response. The decreased P_4–PR function and increased E_2–ERα signalling promote a decline in *ZEB1* expression, a reciprocal increase in miR-200 expression, with further inhibition of ZEB1 and of ZEB2. This phenomenon enables the induction of *OXTR* and *GJA1* gene expression. The increased levels of miR-200s suppress STAT5B, enabling the induction of 20α-HSD, which promotes increased metabolism of P_4 to inactive products within the myometrium, contributing to the decline in PR function. The decreased levels of ZEBs also result in reduced expression of miR-199a-3p and miR-214 and thus increased PTGS2 protein expression, culminating in the synthesis of contractile prostaglandins. Together, these highly orchestrated molecular events enhance myometrial contractility and delivery of the fetus. We propose that similar molecular mechanisms mediate the induction of preterm labour, although the initiating signals (for example, bacterial infection with chorioamnionitis) are different. Taken together, this research has revealed that key miRNAs function as evolutionarily conserved, hormonally controlled modulators of inflammatory and contractile gene expression in the pregnant uterus, with a crucial role in the maintenance of pregnancy and initiation of term and preterm labour.

Review criteria

A PubMed search was performed for articles published between 1971 and 2013, using the keywords "microRNA", "miRNA", "miR", "pregnancy", "parturition", "labour", "progesterone", "progesterone receptor", "estrogen", "estrogen receptor", "inflammation", "inflammatory". Selected original research papers and review articles are discussed in this Review. All articles identified were English language, full-text papers. We also searched the reference lists of identified articles for additional papers.

1. Blencowe, H. *et al.* National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* **379**, 2162–2172 (2012).
2. Peltier, M. R. *et al.* Amniotic fluid and maternal race influence responsiveness of fetal membranes to bacteria. *J. Reprod. Immunol.* **96**, 68–78 (2012).
3. Mendelson, C. R. Minireview: fetal-maternal hormonal signaling in pregnancy and labor. *Mol. Endocrinol.* **23**, 947–954 (2009).
4. Shynlova, O., Lee, Y. H., Srikhajon, K. & Lye, S. J. Physiologic uterine inflammation and labor onset: integration of endocrine and mechanical signals. *Reprod. Sci.* **20**, 154–167 (2013).
5. van Rooij, E., Liu, N. & Olson, E. N. MicroRNAs flex their muscles. *Trends Genet.* **24**, 159–166 (2008).
6. Turner, M. L., Schnorfeil, F. M. & Brocker, T. MicroRNAs regulate dendritic cell differentiation and function. *J. Immunol.* **187**, 3911–3917 (2011).
7. Braun, T. & Gautel, M. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. *Nat. Rev. Mol. Cell Biol.* **12**, 349–361 (2011).
8. Garzon, R., Calin, G. A. & Croce, C. M. MicroRNAs in cancer. *Annu. Rev. Med.* **60**, 167–179 (2009).
9. Mendell, J. T. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* **133**, 217–222 (2008).
10. Boldin, M. P. & Baltimore, D. MicroRNAs, new effectors and regulators of NF-κB. *Immunol. Rev.* **246**, 205–220 (2012).
11. Hawkins, S. M., Buchold, G. M. & Matzuk, M. M. Minireview: The roles of small RNA pathways in reproductive medicine. *Mol. Endocrinol.* **25**, 1257–1279 (2011).
12. Chakrabarty, A. *et al.* MicroRNA regulation of cyclooxygenase-2 during embryo implantation. *Proc. Natl Acad. Sci. USA* **104**, 15144–15149 (2007).
13. Fu, T. Y., Lin, C. T. & Tang, P. C. Steroid hormone-regulated let-7b mediates cell proliferation and basigin expression in the mouse endometrium. *J. Reprod. Dev.* **57**, 627–635 (2011).
14. Hawkins, S. M. *et al.* Dysregulation of uterine signaling pathways in progesterone receptor-Cre knockout of dicer. *Mol. Endocrinol.* **26**, 1552–1566 (2012).
15. Liu, J. L. *et al.* Combined analysis of microRNome and 3′-UTRome reveals a species-specific regulation of progesterone receptor expression in the endometrium of rhesus monkey. *J. Biol. Chem.* **287**, 13899–13910 (2012).
16. Li, R. *et al.* MicroRNA array and microarray evaluation of endometrial receptivity in patients with high serum progesterone levels on the day of hCG administration. *Reprod. Biol. Endocrinol.* **9**, 29 (2011).
17. Su, R. W. *et al.* The integrative analysis of microRNA and mRNA expression in mouse uterus under delayed implantation and activation. *PLoS ONE* **5**, e15513 (2010).
18. Aghajanova, L. & Giudice, L. C. Molecular evidence for differences in endometrium in severe versus mild endometriosis. *Reprod. Sci.* **18**, 229–251 (2011).
19. Xia, H. F. *et al.* Temporal and spatial regulation of miR-320 in the uterus during embryo implantation in the rat. *Int. J. Mol. Sci.* **11**, 719–730 (2010).
20. Xia, H. F. *et al.* Temporal and spatial regulation of let-7a in the uterus during embryo implantation in the rat. *J. Reprod. Dev.* **56**, 73–78 (2010).
21. Hassan, S. S. *et al.* MicroRNA expression profiling of the human uterine cervix after term labor and delivery. *Am. J. Obstet. Gynecol.* **202**, 80 e1–e8 (2010).
22. Nagaraja, A. K. *et al.* Deletion of Dicer in somatic cells of the female reproductive tract causes sterility. *Mol. Endocrinol.* **22**, 2336–2352 (2008).
23. Qian, K. *et al.* Hsa-miR-222 is involved in differentiation of endometrial stromal cells in vitro. *Endocrinology* **150**, 4734–4743 (2009).
24. Gonzalez, G. & Behringer, R. R. Dicer is required for female reproductive tract development and fertility in the mouse. *Mol. Reprod. Dev.* **76**, 678–688 (2009).
25. Montenegro, D. *et al.* Differential expression of microRNAs with progression of gestation and inflammation in the human chorioamniotic membranes. *Am. J. Obstet. Gynecol.* **197**, 289 e1–e6 (2007).
26. Montenegro, D. *et al.* Expression patterns of microRNAs in the chorioamniotic membranes: a role for microRNAs in human pregnancy and parturition. *J. Pathol.* **217**, 113–121 (2009).
27. Hong, X., Luense, L. J., McGinnis, L. K., Nothnick, W. B. & Christenson, L. K. Dicer1 is essential for female fertility and normal development of the female reproductive system. *Endocrinology* **149**, 6207–6212 (2008).
28. Hu, S. J. *et al.* MicroRNA expression and regulation in mouse uterus during embryo implantation. *J. Biol. Chem.* **283**, 23473–23484 (2008).
29. Renthal, N. E. *et al.* miR-200 family and targets, ZEB1 and ZEB2, modulate uterine quiescence and contractility during pregnancy and labor. *Proc. Natl Acad. Sci. USA* **107**, 20828–20833 (2010).
30. Williams, K. C., Renthal, N. E., Condon, J. C., Gerard, R. D. & Mendelson, C. R. MicroRNA-200a serves a key role in the decline of progesterone receptor function leading to term and preterm labor. *Proc. Natl Acad. Sci. USA* **109**, 7529–7534 (2012).
31. Williams, K. C., Renthal, N. E., Gerard, R. D. & Mendelson, C. R. The microRNA (miR)-199a/214 cluster mediates opposing effects of progesterone and estrogen on uterine contractility during pregnancy and labor. *Mol. Endocrinol.* **26**, 1857–1867 (2012).
32. Hardy, D. B., Janowski, B. A., Corey, D. R. & Mendelson, C. R. Progesterone receptor (PR) plays a major anti-inflammatory role in human myometrial cells by antagonism of NF-κB activation of cyclooxygenase 2 expression. *Mol. Endocrinol.* **20**, 2724–2733 (2006).
33. Loudon, J. A., Elliott, C. L., Hills, F. & Bennett, P. R. Progesterone represses interleukin-8 and cyclo-oxygenase-2 in human lower segment fibroblast cells and amnion epithelial cells. *Biol. Reprod.* **69**, 331–337 (2003).
34. Dong, X. *et al.* p54nrb is a transcriptional corepressor of the progesterone receptor that modulates transcription of the labor-associated gene, connexin 43 (*Gja1*). *Mol. Endocrinol.* **23**, 1147–1160 (2009).
35. Shynlova, O., Tsui, P., Dorogin, A. & Lye, S. J. Monocyte chemoattractant protein-1 (CCL-2) integrates mechanical and endocrine signals that mediate term and preterm labor. *J. Immunol.* **181**, 1470–1479 (2008).
36. Cox, S. M., Casey, M. L. & MacDonald, P. C. Accumulation of interleukin-1β and interleukin-6 in amniotic fluid: a sequela of labour at term and preterm. *Hum. Reprod. Update* **3**, 517–527 (1997).
37. Thomson, A. *J. et al.* Leukocytes infiltrate the myometrium during human parturition: further evidence that labour is an inflammatory process. *Hum. Reprod.* **14**, 229–236 (1999).
38. Condon, J. C., Jeyasuria, P., Faust, J. M. & Mendelson, C. R. Surfactant protein secreted by the maturing mouse fetal lung acts as a hormone that signals the initiation of parturition. *Proc. Natl Acad. Sci. USA* **101**, 4978–4983 (2004).
39. Osman, I. *et al.* Leukocyte density and pro-inflammatory cytokine expression in human fetal membranes, decidua, cervix and myometrium before and during labour at term. *Mol. Hum. Reprod.* **9**, 41–45 (2003).
40. Rauk, P. N. & Chiao, J. P. Interleukin-1 stimulates human uterine prostaglandin production through induction of cyclooxygenase-2 expression. *Am. J. Reprod. Immunol.* **43**, 152–159 (2000).
41. Sooranna, S. R. *et al.* Mechanical stretch activates type 2 cyclooxygenase via activator protein-1 transcription factor in human myometrial cells. *Mol. Hum. Reprod.* **10**, 109–113 (2004).
42. Montalbano, A. P., Hawgood, S. & Mendelson, C. R. Mice deficient in surfactant protein A (SP-A) and SP-D or in TLR2 manifest delayed parturition and decreased expression of inflammatory and contractile genes. *Endocrinology* **154**, 483–498 (2013).
43. Shaw, G. & Renfree, M. B. Fetal control of parturition in marsupials. *Reprod. Fertil. Dev.* **13**, 653–659 (2001).
44. Challis, J. R. G., Matthews, S. G., Gibb, W. & Lye, S. J. Endocrine and paracrine regulation of birth at term and preterm. *Endocr. Rev.* **21**, 514–550 (2000).
45. Mitchell, M. D., MacDonald, P. C. & Casey, M. L. Stimulation of prostaglandin E_2 synthesis in human amnion cells maintained in monolayer culture by a substance(s) in amniotic fluid. *Prostaglandins Leukot. Med.* **15**, 399–407 (1984).
46. Romero, R. *et al.* The role of inflammation and infection in preterm birth. *Semin. Reprod. Med.* **25**, 21–39 (2007).
47. Condon, J. C., Hardy, D. B., Kovaric, K. & Mendelson, C. R. Up regulation of the progesterone receptor (PR)-C isoform in laboring myometrium by activation of nuclear factor-κB may contribute to the onset of labor through inhibition of PR function. *Mol. Endocrinol.* **20**, 764–775 (2006).
48. Allport, V. C. *et al.* Human labour is associated with nuclear factor-κB activity which mediates cyclo-oxygenase-2 expression and is involved with the 'functional progesterone withdrawal'. *Mol. Hum. Reprod.* **7**, 581–586 (2001).
49. Lee, Y. S., Terzidou, V., Lindstrom, T., Johnson, M. & Bennett, P. R. The role of CCAAT/enhancer-binding protein β in the transcriptional regulation of COX-2 in human amnion. *Mol. Hum. Reprod.* **11**, 853–858 (2005).
50. Elliott, C. L., Allport, V. C., Loudon, J. A., Wu, G. D. & Bennett, P. R. Nuclear factor-κB is essential for up-regulation of interleukin-8 expression in human amnion and cervical epithelial cells. *Mol. Hum. Reprod.* **7**, 787–790 (2001).
51. Olson, D. M. The role of prostaglandins in the initiation of parturition. *Best Pract. Res. Clin. Obstet. Gynaecol.* **17**, 717–730 (2003).
52. Chow, L. & Lye, S. J. Expression of the gap junction protein connexin-43 is increased in the human myometrium toward term and with the onset of labor. *Am. J. Obstet. Gynecol.* **170**, 788–795 (1994).
53. Fuchs, A. R., Fuchs, F., Husslein, P. & Soloff, M. S. Oxytocin receptors in the human uterus during pregnancy and parturition. *Am. J. Obstet. Gynecol.* **150**, 734–741 (1984).
54. Soloff, M. S., Cook, D. L. Jr, Jeng, Y. J. & Anderson, G. D. In situ analysis of

interleukin-1-induced transcription of cox-2 and il-8 in cultured human myometrial cells. *Endocrinology* **145**, 1248–1254 (2004).
55. Virgo, B. B. & Bellward, G. D. Serum progesterone levels in the pregnant and postpartum laboratory mouse. *Endocrinology* **95**, 1486–1490 (1974).
56. Frydman, R. *et al.* Labor induction in women at term with mifepristone (RU 486): a double-blind, randomized, placebo-controlled study. *Obstet. Gynecol.* **80**, 972–975 (1992).
57. Elliott, C. L., Brennand, J. E. & Calder, A. A. The effects of mifepristone on cervical ripening and labor induction in primigravidae. *Obstet. Gynecol.* **92**, 804–809 (1998).
58. Stenlund, P. M., Ekman, G., Aedo, A. R. & Bygdeman, M. Induction of labor with mifepristone—a randomized, double-blind study versus placebo. *Acta Obstet. Gynecol. Scand.* **78**, 793–798 (1999).
59. Chwalisz, K. The use of progesterone antagonists for cervical ripening and as an adjunct to labour and delivery. *Hum. Reprod.* **9** (Suppl. 1), 131–161 (1994).
60. Pointis, G., Rao, B., Latreille, M. T., Mignot, T. M. & Cedard, L. Progesterone levels in the circulating blood of the ovarian and uterine veins during gestation in the mouse. *Biol. Reprod.* **24**, 801–805 (1981).
61. Shynlova, O., Tsui, P., Jaffer, S. & Lye, S. J. Integration of endocrine and mechanical signals in the regulation of myometrial functions during pregnancy and labour. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **144** (Suppl. 1), 2–10 (2009).
62. Toyoshima, K., Narahara, H., Furukawa, M., Frenkel, R. A. & Johnston, J. M. Platelet-activating factor. Role in fetal lung development and relationship to normal and premature labor. *Clin. Perinatol.* **22**, 263–280 (1995).
63. López Bernal, A., Newman, G. E., Phizackerley, P. J. & Turnbull, A. C. Surfactant stimulates prostaglandin E production in human amnion. *Br. J. Obstet. Gynaecol.* **95**, 1013–1017 (1988).
64. Condon, J. C., Jeyasuria, P., Faust, J. M., Wilson, J. M. & Mendelson, C. R. A decline in progesterone receptor coactivators in the pregnant uterus at term may antagonize progesterone receptor function and contribute to the initiation of parturition. *Proc. Natl Acad. Sci. USA* **100**, 9518–9523 (2003).
65. Mesiano, S. *et al.* Progesterone withdrawal and estrogen activation in human parturition are coordinated by progesterone receptor A expression in the myometrium. *J. Clin. Endocrinol. Metab.* **87**, 2924–2930 (2002).
66. Kalkhoven, E., Wissink, S., van der Saag, P. T. & van der Burg, B. Negative interaction between the RelA(p65) subunit of NF-κB and the progesterone receptor. *J. Biol. Chem.* **271**, 6217–6224 (1996).
67. Hardy, D. B., Janowski, B. A., Chen, C.-C. & Mendelson, C. R. Progesterone receptor inhibits aromatase and inflammatory response pathways in breast cancer cells via ligand-dependent and ligand-independent mechanisms. *Mol. Endocrinol.* **22**, 1812–1824 (2008).
68. Mahendroo, M. S., Cala, K. M. & Russell, D. W. 5α-reduced androgens play a key role in murine parturition. *Mol. Endocrinol.* **10**, 380–392 (1996).
69. Mahendroo, M. S., Porter, A., Russell, D. W. & Word, R. A. The parturition defect in steroid 5α-reductase type 1 knockout mice is due to impaired cervical ripening. *Mol. Endocrinol.* **13**, 981–992 (1999).
70. Andersson, S., Minjarez, D., Yost, N. P. & Word, R. A. Estrogen and progesterone metabolism in the cervix during pregnancy and parturition. *J. Clin. Endocrinol. Metab.* **93**, 2366–2374 (2008).
71. Ishida, M. *et al.* Reproductive phenotypes in mice with targeted disruption of the 20α-hydroxysteroid dehydrogenase gene. *J. Reprod. Dev.* **53**, 499–508 (2007).
72. Piekorz, R. P., Gingras, S., Hoffmeyer, A., Ihle, J. N. & Weinstein, Y. Regulation of progesterone levels during pregnancy and parturition by signal transducer and activator of transcription 5 and 20α-hydroxysteroid dehydrogenase. *Mol. Endocrinol.* **19**, 431–440 (2005).
73. Challis, J. R. Sharp increase in free circulating oestrogens immediately before parturition in sheep. *Nature* **229**, 208 (1971).
74. Buster, J. E. *et al.* Interrelationships of circulating maternal steroid concentrations in third trimester pregnancies. II. C18 and C19 steroids: estradiol, estriol, dehydroepiandrosterone, dehydroepiandrosterone sulfate, δ5-androstenediol, δ androstenedione, testosterone, and dihydrotestosterone. *J. Clin. Endocrinol. Metab.* **48**, 139–142 (1979).
75. Wu, W. X., Myers, D. A. & Nathanielsz, P. W. Changes in estrogen receptor messenger ribonucleic acid in sheep fetal and maternal tissues during late gestation and labor. *Am. J. Obstet. Gynecol.* **172**, 844–850 (1995).
76. Mesiano, S. & Welsh, T. N. Steroid hormone control of myometrial contractility and parturition. *Semin. Cell Dev. Biol.* **18**, 321–331 (2007).
77. Welsh, T. *et al.* Estrogen receptor (ER) expression and function in the pregnant human myometrium: estradiol via ERα activates ERK1/2 signaling in term myometrium. *J. Endocrinol.* **212**, 227–238 (2012).
78. Tibbetts, T. A., Conneely, O. M. & O'Malley, B. W. Progesterone via its receptor antagonizes the pro-inflammatory activity of estrogen in the mouse uterus. *Biol. Reprod.* **60**, 1158–1165 (1999).
79. Murata, T., Narita, K., Honda, K., Matsukawa, S. & Higuchi, T. Differential regulation of estrogen receptor α and β mRNAs in the rat uterus during pregnancy and labor: possible involvement of estrogen receptors in oxytocin receptor regulation. *Endocr. J.* **50**, 579–587 (2003).
80. Piersanti, M. & Lye, S. J. Increase in messenger ribonucleic acid encoding the myometrial gap junction protein, connexin-43, requires protein synthesis and is associated with increased expression of the activator protein-1, c-fos. *Endocrinology* **136**, 3571–3578 (1995).
81. Tsuboi, K. *et al.* Uterine expression of prostaglandin H2 synthase in late pregnancy and during parturition in prostaglandin F receptor-deficient mice. *Endocrinology* **141**, 315–324 (2000).
82. Engstrøm, T. The regulation by ovarian steroids of prostaglandin synthesis and prostaglandin-induced contractility in non-pregnant rat myometrium. Modulating effects of isoproterenol. *J. Endocrinol.* **169**, 33–41 (2001).
83. Albinsson, S. *et al.* MicroRNAs are necessary for vascular smooth muscle growth, differentiation, and function. *Arterioscler. Thromb. Vasc. Biol.* **30**, 1118–1126 (2010).
84. Kim, S. Y. *et al.* miR-143 regulation of prostaglandin-endoperoxidase synthase 2 in the amnion: implications for human parturition at term. *PLoS ONE* **6**, e24131 (2011).
85. Hawkins, S. M. *et al.* Functional microRNA involved in endometriosis. *Mol. Endocrinol.* **25**, 821–832 (2011).
86. Ramon, L. A. *et al.* microRNAs expression in endometriosis and their relation to angiogenic factors. *Hum. Reprod.* **26**, 1082–1090 (2011).
87. Zhao, Z. Z. *et al.* Evaluation of polymorphisms in predicted target sites for micro RNAs differentially expressed in endometriosis. *Mol. Hum. Reprod.* **17**, 92–103 (2011).
88. Burney, R. O. *et al.* MicroRNA expression profiling of eutopic secretory endometrium in women with versus without endometriosis. *Mol. Hum. Reprod.* **15**, 625–631 (2009).
89. Ohlsson Teague, E. M. *et al.* MicroRNA-regulated pathways associated with endometriosis. *Mol. Endocrinol.* **23**, 265–275 (2009).
90. Bracken, C. P. *et al.* A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res.* **68**, 7846–7854 (2008).
91. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
92. Spoelstra, N. S. *et al.* The transcription factor ZEB1 is aberrantly expressed in aggressive uterine cancers. *Cancer Res.* **66**, 3893–3902 (2006).
93. Cochrane, D. R. *et al.* The role of miRNAs in progesterone action. *Mol. Cell. Endocrinol.* **357**, 50–59 (2012).
94. Gregory, P. A. *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* **10**, 593–601 (2008).
95. Postigo, A. A. & Dean, D. C. Differential expression and function of members of the zfh-1 family of zinc finger/homeodomain repressors. *Proc. Natl Acad. Sci. USA* **97**, 6391–6396 (2000).
96. Burk, U. *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep.* **9**, 582–589 (2008).
97. Vandewalle, C., Van Roy, F. & Berx, G. The role of the ZEB family of transcription factors in development and disease. *Cell. Mol. Life Sci.* **66**, 773–787 (2009).
98. Brabletz, S. *et al.* The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. *EMBO J.* **30**, 770–782 (2011).
99. Dudley, D. J., Branch, D. W., Edwin, S. S. & Mitchell, M. D. Induction of preterm birth in mice by RU486. *Biol. Reprod.* **55**, 992–995 (1996).
100. Wu, W. X., Ma, X. H., Zhang, Q. & Nathanielsz, P. W. Characterization of topology-, gestation- and labor-related changes of a cassette of myometrial contraction-associated protein mRNA in the pregnant baboon myometrium. *J. Endocrinol.* **171**, 445–453 (2001).
101. Ou, C. W., Chen, Z. Q., Qi, S. & Lye, S. J. Increased expression of the rat myometrial oxytocin receptor messenger ribonucleic acid during labor requires both mechanical and hormonal signals. *Biol. Reprod.* **59**, 1055–1061 (1998).
102. Ou, C. W., Orsino, A. & Lye, S. J. Expression of connexin-43 and connexin-26 in the rat myometrium during pregnancy and labor is differentially regulated by mechanical and hormonal signals. *Endocrinology* **138**, 5398–5407 (1997).
103. Sparey, C., Robson, S. C., Bailey, J., Lyall, F. & Europe-Finner, G. N. The differential expression of myometrial connexin-43, cyclooxygenase-1 and -2, and Gs α proteins in the upper and lower segments of the human uterus during pregnancy and labor. *J. Clin. Endocrinol. Metab.* **84**, 1705–1710 (1999).
104. Döring, B. *et al.* Ablation of connexin43 in uterine smooth muscle cells of the mouse causes delayed parturition. *J. Cell. Sci.* **119**, 1715–1722 (2006).

105. Nishimori, K. *et al.* Oxytocin is required for nursing but is not essential for parturition or reproductive behavior. *Proc. Natl Acad. Sci. USA* **93**, 11699–11704 (1996).
106. Takayanagi, Y. *et al.* Pervasive social deficits, but normal parturition, in oxytocin receptor-deficient mice. *Proc. Natl Acad. Sci. USA* **102**, 16096–16101 (2005).
107. Puri, C. P. & Garfield, R. E. Changes in hormone levels and gap junctions in the rat uterus during pregnancy and parturition. *Biol. Reprod.* **27**, 967–975 (1982).
108. Power, S. G. & Challis, J. R. The effects of gestational age and intrafetal ACTH administration on the concentration of progesterone in the fetal membranes, endometrium, and myometrium of pregnant sheep. *Can. J. Physiol. Pharmacol.* **65**, 136–140 (1987).
109. Csapo, A. I., Eskola, J. & Tarro, S. Gestational changes in the progesterone and prostaglandin F levels of the guinea-pig. *Prostaglandins* **21**, 53–64 (1981).
110. Runnebaum, B. & Zander, J. Progesterone and 20α-dihydroprogesterone in human myometrium during pregnancy. *Acta Endocrinol. Suppl. (Copenh.)* **150**, 3–45 (1971).
111. Penning, T. M. & Drury, J. E. Human aldo-keto reductases: Function, gene regulation, and single nucleotide polymorphisms. *Arch. Biochem. Biophys.* **464**, 241–250 (2007).
112. Richer, J. K. *et al.* Convergence of progesterone with growth factor and cytokine signaling in breast cancer. Progesterone receptors regulate signal transducers and activators of transcription expression and activity. *J. Biol. Chem.* **273**, 31317–31326 (1998).
113. Stocco, C. O., Chedrese, J. & Deis, R. P. Luteal expression of cytochrome P450 side-chain cleavage, steroidogenic acute regulatory protein, 3β-hydroxysteroid dehydrogenase, and 20α-hydroxysteroid dehydrogenase genes in late pregnant rats: effect of luteinizing hormone and RU486. *Biol. Reprod.* **65**, 1114–1119 (2001).
114. Chen, J. & Nathans, J. Estrogen-related receptor β/NR3B2 controls epithelial cell fate and endolymph production by the stria vascularis. *Dev. Cell* **13**, 325–337 (2007).
115. Yin, G. *et al.* TWISTing stemness, inflammation and proliferation of epithelial ovarian cancer cells through MIR199A2/214. *Oncogene* **29**, 3545–3553 (2010).
116. Loebel, D. A., Tsoi, B., Wong, N. & Tam, P. P. A conserved noncoding intronic transcript at the mouse *Dnm3* locus. *Genomics* **85**, 782–789 (2005).
117. Olson, D. M. *et al.* Myometrial activation and preterm labour: evidence supporting a role for the prostaglandin F receptor—a review. *Placenta* **24** (Suppl. A), 47–54 (2003).
118. Havelock, J. C. *et al.* Human myometrial gene expression before and during parturition. *Biol. Reprod.* **72**, 707–719 (2005).
119. Menon, R. *et al.* Biomarkers of spontaneous preterm birth: an overview of the literature in the last four decades. *Reprod. Sci.* **18**, 1046–1070 (2011).
120. Wittmann, J. & Jäck, H. M. Serum microRNAs as powerful cancer biomarkers. *Biochim. Biophys. Acta* **1806**, 200–207 (2010).
121. Etheridge, A., Lee, I., Hood, L., Galas, D. & Wang, K. Extracellular microRNA: a new source of biomarkers. *Mutat. Res.* **717**, 85–90 (2011).
122. Frost, R. J. & Olson, E. N. Control of glucose homeostasis and insulin sensitivity by the Let-7 family of microRNAs. *Proc. Natl Acad. Sci. USA* **108**, 21075–21080 (2011).
123. van Rooij, E., Marshall, W. S. & Olson, E. N. Toward microRNA-based therapeutics for heart disease: the sense in antisense. *Circ. Res.* **103**, 919–928 (2008).
124. Montgomery, R. L. *et al.* Therapeutic inhibition of miR-208a improves cardiac function and survival during heart failure. *Circulation* **124**, 1537–1547 (2011).
125. Obad, S. *et al.* Silencing of microRNA families by seed-targeting tiny LNAs. *Nat. Genet.* **43**, 371–378 (2011).
126. Small, E. M. & Olson, E. N. Pervasive roles of microRNAs in cardiovascular biology. *Nature* **469**, 336–342 (2011).
127. Concepcion, C. P., Bonetti, C. & Ventura, A. The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer J.* **18**, 262–267 (2012).
128. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
129. Ruvkun, G. The perfect storm of tiny RNAs. *Nat. Med.* **14**, 1041–1045 (2008).

Acknowledgements
The authors' research was supported by the National Institutes of Health (NIH 5-P01-HD011149) and the March of Dimes Foundation (Prematurity Research Grant No. 21-FY11-30).

Author contributions
The authors contributed equally to all aspects of this article.

Glotov et al. BMC Systems Biology 2015, 9(Suppl 2):S4
http://www.biomedcentral.com/1752-0509/9/S2/S4

BMC Systems Biology

RESEARCH Open Access

Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome)

Andrey S Glotov[1,2], Evgeny S Tiys[3,4], Elena S Vashukova[1], Vladimir S Pakin[1], Pavel S Demenkov[3,4], Olga V Saik[3,4], Timofey V Ivanisenko[3,4], Olga N Arzhanova[1], Elena V Mozgovaya[1], Marina S Zainulina[1], Nikolay A Kolchanov[3,4], Vladislav S Baranov[1,2], Vladimir A Ivanisenko[3,4*]

From IX International Conference on the Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS\SB-2014)
Novosibirsk, Russia. 23-28 June 2014

Abstract

Background: Pre-eclampsia is the most common complication occurring during pregnancy. In the majority of cases, it is concurrent with other pathologies in a comorbid manner (frequent co-occurrences in patients), such as diabetes mellitus, gestational diabetes and obesity. Providing bronchial asthma, pulmonary tuberculosis, certain neurodegenerative diseases and cancers as examples, we have shown previously that pairs of inversely comorbid pathologies (rare co-occurrences in patients) are more closely related to each other at the molecular genetic level compared with randomly generated pairs of diseases. Data in the literature concerning the causes of pre-eclampsia are abundant. However, the key mechanisms triggering this disease that are initiated by other pathological processes are thus far unknown. The aim of this work was to analyse the characteristic features of genetic networks that describe interactions between comorbid diseases, using pre-eclampsia as a case in point.

Results: The use of ANDSystem, Pathway Studio and STRING computer tools based on text-mining and database-mining approaches allowed us to reconstruct associative networks, representing molecular genetic interactions between genes, associated concurrently with comorbid disease pairs, including pre-eclampsia, diabetes mellitus, gestational diabetes and obesity. It was found that these associative networks statistically differed in the number of genes and interactions between them from those built for randomly chosen pairs of diseases. The associative network connecting all four diseases was composed of 16 genes (*PLAT, ADIPOQ, ADRB3, LEPR, HP, TGFB1, TNFA, INS, CRP, CSRP1, IGFBP1, MBL2, ACE, ESR1, SHBG, ADA*). Such an analysis allowed us to reveal differential gene risk factors for these diseases, and to propose certain, most probable, theoretical mechanisms of pre-eclampsia development in pregnant women. The mechanisms may include the following pathways: [TGFB1 or TNFA]-[IL1B]-[pre-eclampsia]; [TNFA or INS]-[NOS3]-[pre-eclampsia]; [INS]-[HSPA4 or CLU]-[pre-eclampsia]; [ACE]-[MTHFR]-[pre-eclampsia].

Conclusions: For pre-eclampsia, diabetes mellitus, gestational diabetes and obesity, we showed that the size and connectivity of the associative molecular genetic networks, which describe interactions between comorbid diseases, statistically exceeded the size and connectivity of those built for randomly chosen pairs of diseases. Recently, we have shown a similar result for inversely comorbid diseases. This suggests that comorbid and inversely comorbid diseases have common features concerning structural organization of associative molecular genetic networks.

* Correspondence: salix@bionet.nsc.ru
[3]The Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
Full list of author information is available at the end of the article

Background

Pre-eclampsia (PE) is the leading cause of maternal and foetal morbidity and mortality. It is a pregnancy complication, predominantly occurring after 20-weeks of gestation, as well as in labour, and it is characterized by multiple organ dysfunction syndromes, including the dysfunction of the kidneys, liver, vascular and nervous systems, and the foetoplacental complex [1,2]. The general clinical symptoms of PE are oedema, proteinuria and hypertension. The clinical outcome of PE may not always be predictable. Either form of PE can be extremely insidious, rapidly progressing, and, even in the absence of one of its general symptoms, may lead to life threatening complications for the mother and foetus [3]. In 70-80% of cases, PE is secondary to an underlying disease [1]. Pre-eclampsia risk factors include cardiovascular diseases (arterial hypertension), kidney, liver and gastrointestinal tract diseases, endocrine disorders (obesity, diabetes mellitus), and autoimmune diseases (antiphospholipid syndrome) [1,3,4]. According to meta-analysis data, women with a history of PE have 1.79 times the risk of venous thromboembolism, 1.81 times the risk of stroke, 2.16 times the risk of ischemic heart disease and 3.7 times the risk of hypertensive disease, compared with women without PE [5]. Thus far, it remains unclear whether the presence of pathological processes before pregnancy predisposes one to PE, or whether defects in multiple organs and systems, induced by PE, are responsible for the development of extragenital diseases in the future. Such joint manifestations of diseases are called comorbidities [6] or syntropies [7]. Likewise, inversely comorbid [8] or dystropic [9] diseases statistically rare co-occur in patients as compared with co-occurrence that can be expected by chance. Previously, for asthma, tuberculosis, certain cancers and neurodegenerative diseases, we have shown that inversely comorbid diseases are more closely related to each other at the molecular level in comparison with randomly chosen pairs of diseases [10].

In recent years, bioinformatics methods have been widely used for modelling different pathological processes, analysing the molecular mechanisms of their development, identifying possible markers, and systematizing available data. Ample evidence regarding the influence of genetics on comorbidities has accumulated in the literature. Computer-based, text-mining methods were developed for efficient extraction of knowledge from the scientific literature. At the present time, COREMINE and MeSHOPs, which analyse the co-occurrence of biomedical terms [11,12], and STRING [13] and the MedScan system, which are based on the parsing of natural language texts [14], are widely used.

We have developed the ANDSystem, which was designed for the automated extraction of knowledge from natural language texts regarding the properties of molecular biological objects and their interactions in living systems [15]. Using this system, we have reconstructed protein-protein networks for proteins that are associated with water-salt metabolism and sodium deposition processes in healthy volunteers [16], as well as protein-protein interaction networks for *Helicobacter pylori*, which are associated with the functional divergence of *H. pylori*, isolated from patients with early gastric cancer [17]. We have also reconstructed associative networks representing molecular genetic interactions between proteins, genes, metabolites and molecular processes associated with myopia and glaucoma [18], and with cardiovascular diseases [19].

In the current study, we used the ANDSystem for the reconstruction of associative networks (the preeclampsia associome) representing molecular genetic interactions between genes associated with PE, diabetes mellitus (DM), gestational diabetes (GD) and obesity (Ob). We conducted an analysis of these networks to reveal differential and common risk factors for these diseases.

Finding pathways common to the indicated multifactorial diseases would contribute to a better understanding of the characteristic features of pre-eclampsia pathogenesis, as well as to the development of new diagnostic, preventative and therapeutic methods.

Results

Pre-eclampsia: its association, via comorbid genes, with diabetes mellitus, obesity and gestational diabetes

The main goal of the current study was to identify comorbid genes whose dysfunction or mutation represent common risk factors for diseases that are concurrent with PE. To this end, we relied on published data [3,4] regarding the four most significant and widespread pathologies concurrent with PE: DM, Ob, GD and pyelonephritis. Furthermore, using the ANDSystem and Pathway Studio software, we reconstructed associative networks (disease-protein/gene-disease) comprising interactions between the above diseases via their associated genes. Subsequently, reduction was achieved by eliminating pyelonephritis, as genes associated with nephritis were not associated with PE and the other analysed disorders. Using the ANDSystem, we identified 1,051 proteins/genes associated with PE, Ob, DM and/or GD. Using Pathway Studio, 1,138 proteins/genes were identified. The results of both programs were in good agreement regarding the number of genes in groups associated with particular diseases (Figure 1). Unfortunately, we were not able to use STRING for the reconstruction of such networks, as this program does not provide information about protein/gene-disease associations.

The number of proteins/genes common to different combinations of the examined diseases is shown in Figure 1. We assumed that comorbid diseases are more

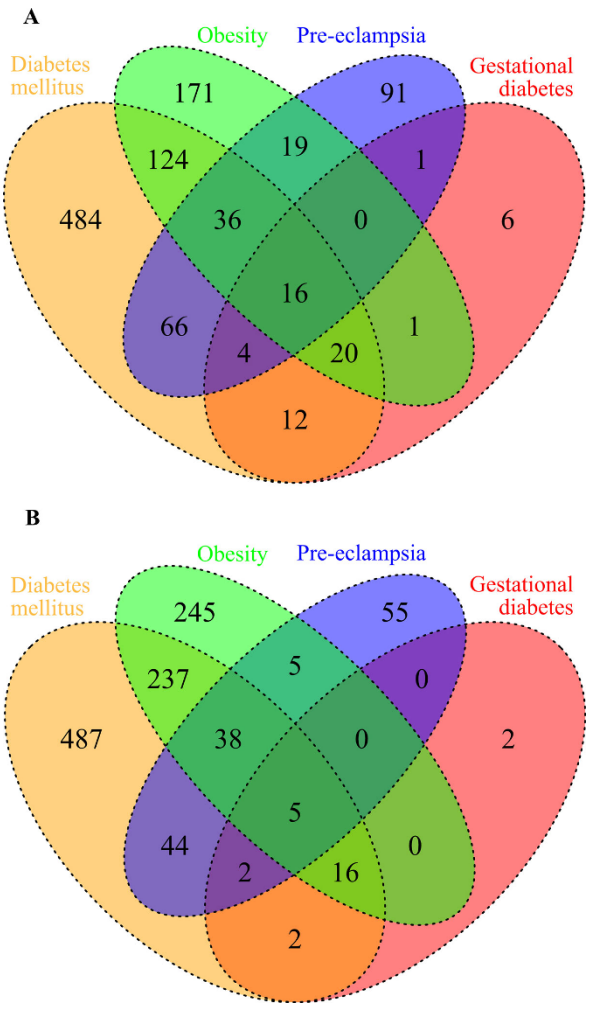

Figure 1 Venn diagram demonstrating the intersections of the lists of genes associated with the analysed diseases (PE, DM, GD, Ob) according to the ANDSystem (A) and Pathway Studio (B).

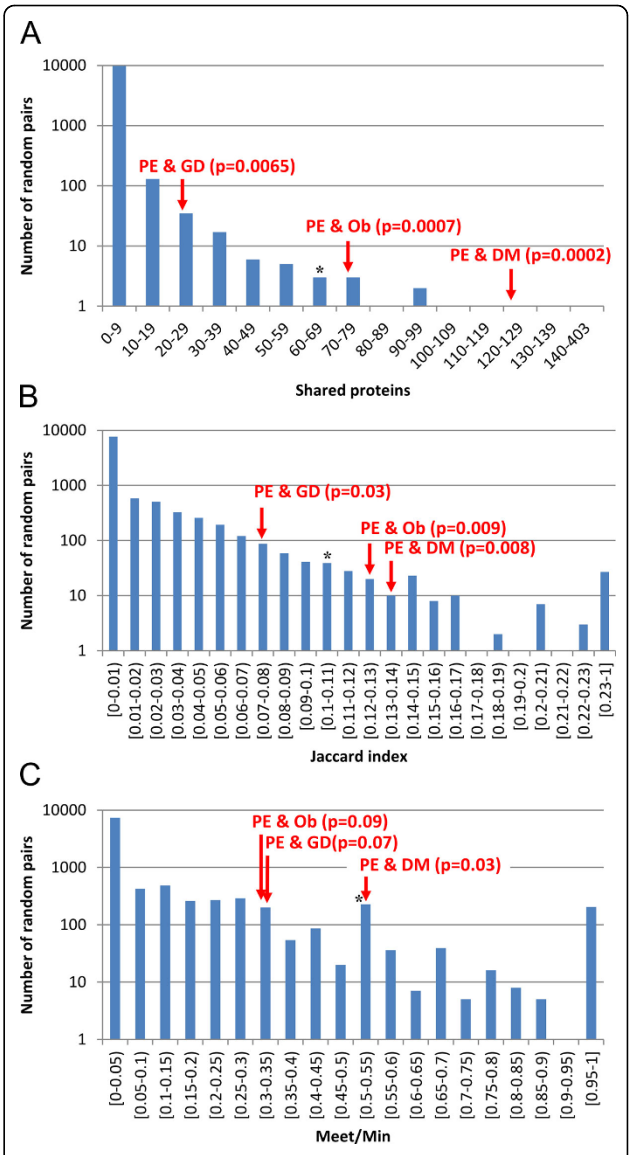

Figure 2 Comparison of analysed and random networks by intersection (A), Jaccard (B) and Meet/Min (C) indices. Bars show the distribution of the value for the features of the associative networks for randomly chosen disease pairs. Arrows indicate PE & GD, PE & DM and PE & Ob comorbid disease pairs. Asterisks indicate the position of inversely comorbid disease pairs (bronchial asthma and pulmonary tuberculosis) [10].

closely interrelated, via the common proteins/genes associated with them, as compared with randomly chosen disease pairs. To test this assumption, we calculated the distribution of three relation indices of random disease-protein/gene-disease networks built for random disease pairs: I_{AB} (number of shared proteins), J_{AB} (Jaccard index) and M_{AB} (Meet/Min). All three disease pairs (PE & DM, PE & GD, PE & Ob) were significantly connected by the I_{AB} and J_{AB} indices at $p < 0.05$ (Figure 2). Only PE & DM pair was significantly different by M_{AB} index ($p < 0.05$) from randomly generated pairs of diseases. Thus, PE and DM were found to be the most significantly associated disease pair for all three relation indices.

Next, we tested the hypothesis whether comorbid proteins/genes common to comorbid disease pairs interact more closely compared to a set of randomly chosen proteins/genes. Comparison of the associative molecular genetic networks with random ones demonstrated that the networks that describe the interactions between the comorbid proteins/genes for all three disease pairs (PE & GD, PE & DM, and PE & Ob) exhibited significantly greater connectivities than those of the random networks ($p < 0.001$).

Of particular interest was an appended analysis of the associative molecular genetic networks built for proteins associated concurrently with four comorbid diseases (PE, DM, Ob and GD). The three programs used to

build this network were the ANDSystem, Pathway Studio and STRING (Figure 3). As Figure 3A shows, the ANDSystem network comprised 32 objects: 16 proteins and 16 genes, as well as 142 interactions. The ANDSystem has an advantageous feature: an object pair can also be associated concurrently with links of several types. For this reason, the number of associated object pairs, 87, was smaller than the number of links. The ANDSystem

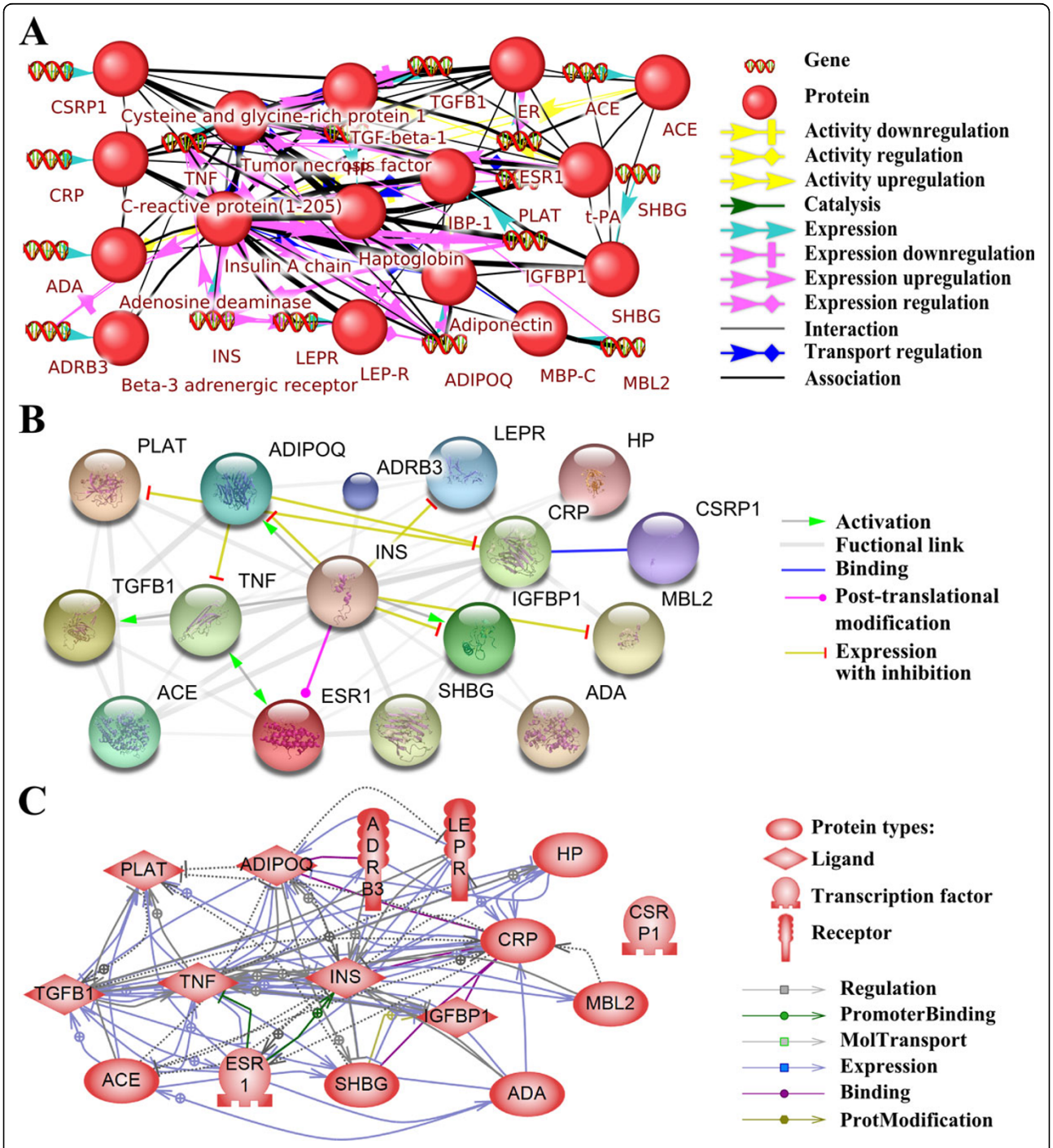

Figure 3 ANDSystem (A), STRING (B) and Pathway Studio (C) networks describing shared proteins/genes. Interactions between proteins/genes concurrently shared in PE, DM, Ob and GD are shown. In the ANDSystem network, proteins are shown as balls, and patterns with double helices designate genes. In the STRING network, proteins/genes are shown as balls (gene names are given). The colour and shape of the arrows indicate the type of association between the objects in all the networks.

represented cases of the regulation of protein activity (six links), including up-regulation (two links) and down-regulation (three links) of protein activity; gene expression regulation (37 links), including up-regulation (seven links) and down-regulation (seven links); protein-protein interactions (two links); protein transport regulation (10 links); catalysis (one link); expression (16 links) and association (70 links). To compare the ANDSystem network with those of the STRING and Pathway Studio, the ANDSystem network was transformed into a protein-protein interaction network, with links from the genes assigned to their respective proteins, while links from genes as separate vertices were deleted from the network. Such a network contained 45 interconnected protein pairs.

The STRING network (Figure 3B) contained 16 proteins/genes, and 45 gene pairs connected by 47 links, including five different types: activation (four links), expression with inhibition (seven links), binding (one link), post-translational modification (one link), and functional links (34 links). The functional links in STRING were determined on the basis of Neighbourhood in the Genome, Gene Fusions, Co-occurrence Across Genomes, Co-Expression, Experimental/Biochemical Data, Association in Curated Databases, and Co-Mentioned in PubMed Abstracts [13].

The network built by Pathway Studio (Figure 3C) contained 16 proteins/genes, and 62 pairs of genes connected by 98 links, including six different types: binding (five links), expression (55 links), molecular transport (19 links), promoter binding (two links), protein modification (one link) and regulation (16 links).

There was a significant difference between the comorbid and random networks ($p < 0.001$), not only for disease pairs, but also for the associative molecular genetic networks that describe the interactions between proteins/genes associated concurrently with all four diseases (PE, DM, GD, Ob) (Figure 3A). These results demonstrated that comorbid proteins/genes are presumably involved in shared biological processes. This can explain the increased number of interactions between proteins/genes, as compared with those for associative molecular genetic networks of randomly chosen proteins/genes. Confirmation of this hypothesis would shed light on the molecular mechanisms underlying the interactions between comorbid diseases.

Analysis of overrepresentation of Gene Ontology (GO) processes

Overrepresentation of GO biological processes was analysed for the group of proteins/genes associated with single diseases (PE, DM, GD and Ob) and pairs of diseases (PE & DM, PE & GD, PE & Ob), as well as concurrently with all four diseases. In each of these cases, more than 1,000 overrepresented processes were found (Additional file 1). Among these were a high number of quite general processes for which thousands of genes have been annotated. The connectivity rate (CR) was calculated for each process listed in Additional file 1 to check how closely the proteins/genes, which caused an overrepresentation of processes, interacted. After ranking the overrepresented biological processes according to the CRs, 313 processes had the highest CR (equal to 1) (see Additional file 1). Just as expected, generalized, nonspecific biological processes had smaller CR values in the majority of cases as compared with specialized processes involving a relatively small number of genes.

Among the overrepresented biological processes with a maximum CR were positive regulation of monooxygenase activity, regulation of fat cell differentiation, regulation of lipid metabolic process, nitric oxide and carbon monoxide metabolism, regulation of protein kinase B signalling cascade, regulation of NF-kappa B transcription factor activity, regulation of glucose metabolism and transport, regulation of cellular response to oxidative stress, regulation of cytokine production, regulation of cell cycle process and others. Thus, the use of the CR index in the GO enrichment analysis revealed the specific GO processes and lower the rank of less informative general processes.

Reconstruction of associative pathways describing potential molecular mechanisms via comorbid genes involved in overrepresented GO biological processes

The next step of the current study was to reconstruct the molecular pathways connecting PE with DM, Ob, and GD, via interactions between the specific and comorbid genes. The Pathway Discovery module of the ANDVisio software was used to trace separate pathways in the network of molecular genetic interactions associated concurrently with all four pathologies. The Pathway Discovery module was used to search for pathways in the network using patterns set by the user.

The patterns were of the following type: <PE> - <any protein/gene specific to PE> - <any comorbid protein/gene> - <any protein/gene specific to Ob or GD, or DM> - <Ob or GD, or DM>. The program chose all the pathways meeting the pattern's criteria: the starting link was PE; the second link of the chain should be one of the proteins/genes associated with PE, exceptions were proteins/genes comorbid for all four diseases (4-comorbid); the third link should be one of the 4-comorbid proteins/genes (PLAT, ADIPOQ, ADRB3, LEPR, HP, TGFB1, TNFA, INS, CRP, CSRP1, IGFBP1, MBL2, ACE, ESR1, SHBG, ADA); the fourth link should be one of the proteins specific to Ob, GD, or DM, with the exception of 4-comorbid proteins/genes. The last link should be one of the diseases (Ob, GD or DM). The total number of identified pathways was more than 50. These were combined into a single pathway network (Figure 4).

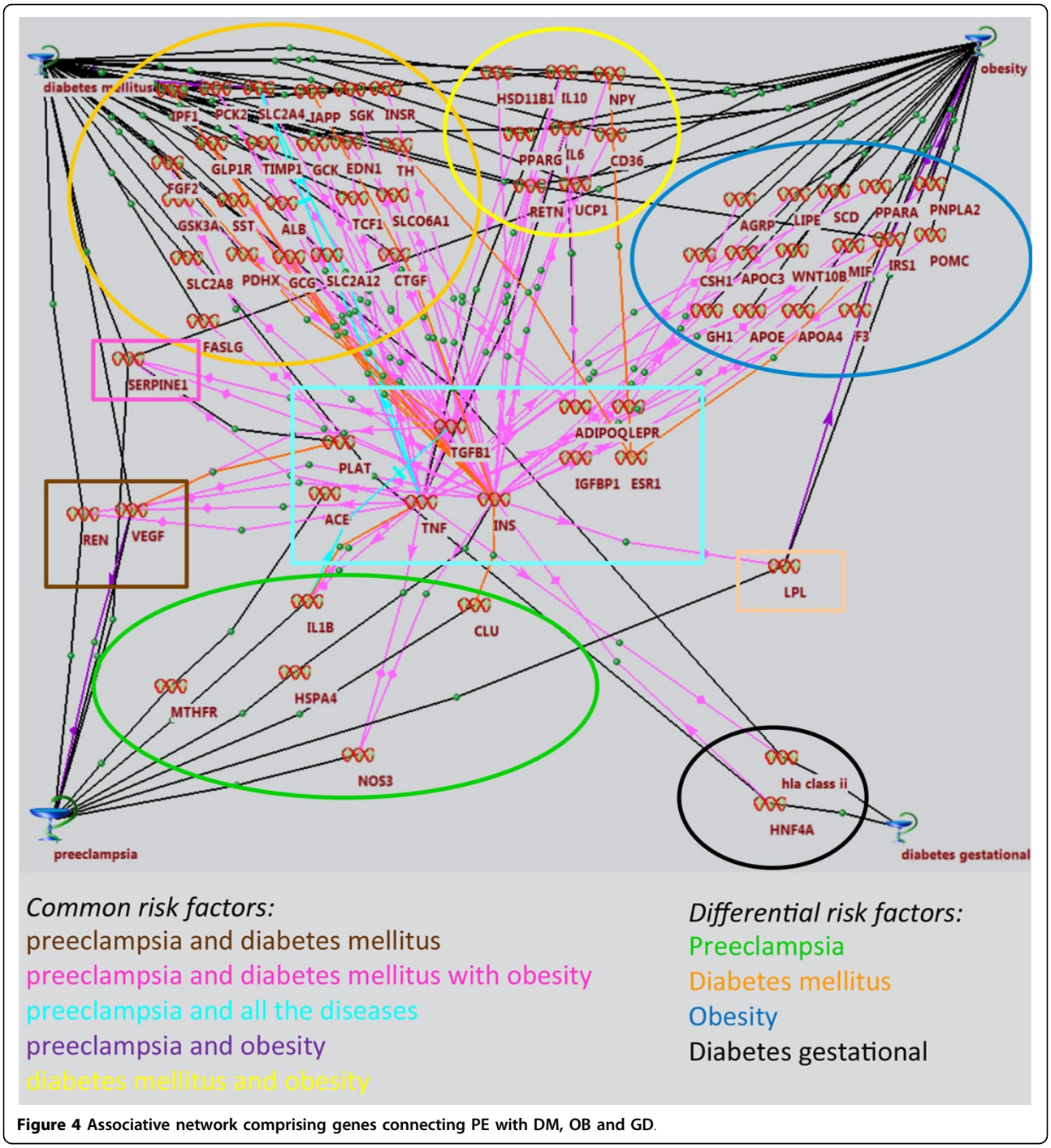

Figure 4 Associative network comprising genes connecting PE with DM, OB and GD.

Common, as well as specific, risk factors were distinguished for the following combinations of diseases: PE and DM; PE and Ob; PE and DM, Ob; PE and DM, Ob, GD; (see Figure 4). The largest number of connections was obtained for the *TNFA, TGFB1* and *INS* genes, which revealed specialized GO processes with maximum CRs, such as: «positive regulation of protein kinase B signalling», «cascade regulation of NF-kappa B transcription factor activity», «regulation of mitosis», «regulation of nuclear division», «regulation of protein secretion MAPK cascade», «positive regulation of protein transport», «regulation of protein complex assembly», «positive regulation of cell migration», «positive regulation of secretion», «positive regulation of cellular component movement», «positive regulation of organelle organization», «regulation of mitotic cell cycle»,

«regulation of immune effector process», «intracellular protein kinase cascade», «regulation of cellular component biogenesis», «regulation of cell cycle process», «regulation of organelle organization», «regulation of cell cycle» (see Additional file 1).

An associative pathway network connecting PE, via the *PLAT, ADIPOQ, LEPR, TGFB1, TNFA, INS, IGFBP1, ACE* and *ESR1* genes, with DM, OB and GD incorporated 66 genes with 167 connections (see Figure 4). Most of these connections (78) corresponded to the "association" type (shown in black). Sixty-nine of them could be referred to "expression regulation" types and 13 as "co-expression" (shown in red); eight comprised "down regulation", "degradation regulation", and "degradation downregulation" (shown in violet).

The differential network of PE risk factors included seven genes (interleukin-1-beta (IL1B), endothelial (NOS3) NO-synthase, heat shock 70 kDa protein 4 (HSPA4), apolipoprotein J (CLU) and 5,10-methylenetetrahydrofolate reductase (MTHFR).

Thus, whereas all the identified PE risk factors might be treated as potential markers of this disease, the most probable molecular mechanism underlying PE, DM, OB and GD includes the pathway starting from the *TGFB1, TNFA, INS* and *ACE* genes, through the *IL1B, NOS3, HSPA4* (HSP74), *CLU* and *MTHFR* genes, and eventually to PE.

Thus, the probable chains of molecular events on the way to combined PE, in this context, are as follows: TGFB1 or TNFA - IL1B - PE; TNFA or INS - NOS3 - PE; INS - HSPA4 or CLU - PE; ACE - MTHFR - PE.

Discussion

The associative networks analysed in this work (see Figures 1, 2, 3, 4) appeared to be significant for the understanding of the nature of PE, thereby supporting the hypothesis that PE represents a stable complex of clinical manifestations [1,3,4]. The key players in the reconstructed networks are comorbid genes which, on the one hand, contribute to the development of PE and its pathogenically related disorders, and, on the other hand, may play the role of "triggers" in the presence of other pre-eclampsia-promoting factors (genes and proteins). Comorbid genes are characteristic of many multifactorial diseases [20]. Moreover, many comorbid diseases may involve various pathophysiological mechanisms [20], and the construction of associative networks makes it possible to understand their molecular interrelations.

An analysis of reconstructed associative networks, which describe interactions between comorbid proteins/genes associated with different pair combinations of PE with DM, Ob, and GD, demonstrated that comorbid diseases differ in a statistically significant manner from random disease pairs. The differences concern both the number of common genes associated with the diseases and the interactions between such genes. The number of vertices in the comorbid networks, as well as the number of interactions between the vertices, exceeded those of random disease pairs. At the same time, the density of connections in the associative molecular genetic network describing the interactions between proteins/genes associated concurrently with all four diseases also differed significantly from those of the random networks formed by random sets of proteins/genes. Interestingly, we also observed the same regularity for inversely comorbid diseases [10]. It has been shown that the associative networks reconstructed for pairs of inversely comorbid diseases, including bronchial asthma and pulmonary tuberculosis, as well as nine pairs formed by neurodegenerative (Parkinson disease, schizophrenia, Alzheimer disease) and cancer diseases (colorectal neoplasms, prostatic neoplasms, lung neoplasms), significantly differed from the networks that describe interactions between random diseases. An example of the mutual arrangement of inversely comorbid (bronchial asthma and pulmonary tuberculosis) and comorbid diseases is shown in Figure 2.

Our current results are in many respects consistent with those of epidemiological studies worldwide. It has been amply demonstrated that the common risk factors of PE were DM, Ob and GD [1,2,21-26]. In most studies, DM is a leading risk factor, as it occurs in more than half of the women with PE [1,2,24]. Furthermore, DM is more strongly associated with a late-onset of the disease, which prevails among all the cases [24,25]. A study of twin gestations supports our reasoning. In this study, an evaluation of associated factors in PE gestations and a comparison of the incidence of pregnancy complications among twins with and without PE demonstrated that a high pregnancy body mass index (BMI) and diabetes were associated with PE [27].

We identified 16 genes encoding shared proteins in the molecular network, built using the literature- and database-mining (ANDSystem, Pathway Studio and STRING), that simultaneously connected with PE, DM, GD and Ob. Most shared genes determined in this study encode proteins controlling energy metabolism, and are associated with the immune response and inflammation.

An analysis of the associations of these genes with PE and DM, GD and Ob obtained in case-control, family-based, and meta-analyses studies, which we conducted using the HuGE Navigator, revealed that 14 of the 16 shared genes were associated with at least one of the diseases (see Table 1). Two genes (*CSRP1* and *PLAT*) had never been shown to be associated with PE and DM, GD and Ob. Four shared genes (*ACE, ADIPOQ,*

Table 1. Statistics of gene-disease associations for PE, DM, GD and Ob obtained with the HuGE Navigator

Gene name	PE	DM	GD	Ob
ACE	39	244	2	77
ADA	-	6	-	-
ADIPOQ	4	156	4	176
ADRB3	1	49	4	145
CRP	2	20	-	28
CSRP1	-	-	-	-
ESR1	7	21	-	36
HP	2	36	-	5
IGFBP1	-	7	-	5
INS	1	88	4	26
LEPR	7	35	1	154
MBL2	4	14	1	1
PLAT	2	3	-	1
SHBG	-	6	1	7
TGFB1	8	33	-	8
TNFA	24	132	5	83

The number of associations determined by case-control, family-based and meta-analysis studies are shown.

MBL2, TNFA) were found to be associated with all the diseases.

We believe that the identification of these genes in the current study is of importance because they encode proteins important for the development of diseases, as confirmed by experimental studies (Table 1).

Angiotensin-converting enzyme (ACE) plays a key role in regulating blood pressure by influencing vascular tone by activating the vasoconstrictor angiotensin II and inactivating the vasodilatory peptide bradykinin. Inter-individual differences in blood ACE levels are at least in part explained by the presence of an insertion/deletion (I/D) polymorphism in intron 16 of the ACE gene, with higher ACE levels observed in D allele carriers. The results of many studies confirmed the association of ACE polymorphism with PE [28]. Other studies have indicated that the ACE gene is a factor that contributes to the manifestation of GD [29], diabetic nephropathy and Ob [30,31].

It has been shown that polymorphisms in the adiponectin gene (*ADIPOQ*) modulate the circulating concentration of adiponectin. Abnormal adiponectin levels, as well as *ADIPOQ* polymorphisms, have been associated with PE [32]. Some variants of this gene are associated with the occurrence of GD [33], while other polymorphisms may contribute to type 2 DM risk [34] and Ob in adults [35].

Mannose-binding lectin (MBL) is involved in the maintenance of an inflammatory environment in the uterus. High MBL levels have been associated with successful pregnancies, whereas low levels are involved in PE development. Association between polymorphisms in the structural and promoter regions of the *MBL2* gene and PE have been evaluated [36]. MBL gene polymorphisms are associated with GD and with type 2 DM [37,38]; in addition, MBL deficiency may confer a risk of Ob and insulin resistance [39].

Tumour necrosis factor-alpha (TNF-α) participates in the immune response and inflammation. Many studies have showed that there is an association between the *TNFA* gene and PE among Europeans [2,40]. The -308 G−>A polymorphism of the *TNFA* promoter gene is involved in the pathophysiology of insulin resistance and GD [41]. The same polymorphism is a genetic risk factor for the development of type 2 DM [42]. Individuals who carry the -308A *TNFA* gene variant have a 23% greater risk of developing obesity compared with controls, and they showed significantly higher systolic arterial blood pressure and plasma insulin levels, supporting the hypothesis that the *TNFA* gene is involved in the pathogenesis of the metabolic syndrome [43].

The PE associome contains more links than each of the individual networks. The identified, shared genes have been classified according to GO. Such a network was needed for a GO overrepresentation analysis. The presence of processes identified by the GO analysis in the pathogenesis of PE is not surprising. The central hypothesis of our understanding of PE is that it results from ischaemia of the placenta, which in turn releases factors into the maternal circulation that are capable of inducing the clinical manifestations of the disease [2]. Multiple pathogenetic mechanisms have been implicated in this disorder, including an imbalance between angiogenic and anti-angiogenic factors, autoantibodies to the type-1 angiotensin II receptor, platelet and thrombin activation, defective deep placentation, intravascular inflammation, endothelial cell activation and/or dysfunction, and oxidative and endoplasmic reticulum stress that promote the differentiation of trophoblasts from a proliferative to an invasive phenotype, regulate cell homeostasis through their involvement in post-translational modifications and protein folding, and induce the release of proinflammatory cytokines and chemokines. Other mechanisms include hypoxia and trophoblast invasion, which down-regulate the expression of transforming growth factor β3 (TGF-β3) and hypoxia-inducible factors (HIF-1α and HIF-2α) [2,44]. These results indicated the contribution of common, non-specific, pathological processes to the development of PE, DG, GD and Ob.

In addition to the identification of common proteins/genes associated with different pathological processes, another goal of the study was to find unique markers for PE. To do so, we reconstructed potential mechanisms of molecular interactions using the ANDSystem software, a program that allows the identification of the largest number of links (see Figure 4). Although the central network

core of these pathways contained only nine common genes (*PLAT, ADIPOQ, LEPR, TGFB1, TNFA, INS, IGFBP1, ACE, ESR1*), it incorporated 68 genes with 174 connections between them, and differential factor risks of PE were identified: the *IL1B, NOS3, HSPA4, CLU* and *MTHFR* genes. The contributions of many of these genes to the pathogenesis of PE has been confirmed by numerous studies [2,45-50]. Here, we showed for the first time that these genes can be specifically involved in the pathogenesis of PE. However, it is not yet clear why these genes have a greater involvement in PE. The possible trigger mechanisms of combined PE are linked to the processes that are carried out by the products of the identified genes, namely, inflammation (*IL1B*), endothelial dysfunction (*NOS3*), heat shock and stress (*HSPA4*), stabilizing cell membranes at diverse fluid-tissue interfaces and protecting the vascular endothelium from an attack by some factors in plasma, such as active complement complexes (*CLU*), and homocysteine metabolism (*MTHFR*).

In addition, the results are of particular importance in regard to the theory of confounding assumptions as false mechanisms of genetic association when the factor is associated with a confound, but not the phenotype, and a confound, in turn, is associated with the phenotype [51,52]. The identified genes can act as such a confound.

Conclusions

The current results broaden our knowledge of the molecular mechanisms of the interactions between comorbid diseases. This reconstruction of associative molecular genetic networks that describe interactions between PE and comorbid diseases (GD, Ob, and DM) differed significantly from partner networks built for random disease pairs. Networks between PE and comorbid diseases had a larger number of genes and links between them. With this in mind, it is of interest that similar features of associative network structure have been observed for inversely comorbid diseases [10]. It can be suggested that comorbid and inversely comorbid relationships between diseases involve larger sets of closely interrelated genes larger than those for random pairs of diseases. In the future, we intend to perform a scale analysis that connects different disease pairs to detect potential comorbid/inversely comorbid diseases for all the possible disease pairs via which these diseases can interact. Reconstruction and analysis of the PE associome is useful for revealing the genetic factors involved in the pathogenesis of the disease and for identifying its differential risk factors, as well as for modelling the theoretical mechanisms of PE development in pregnant women with underlying diseases, such as DB, Ob or GD.

Methods

We used three systems that allowed the automated reconstruction of networks that describe the interactions between proteins/genes and diseases: STRING [13], Pathway Studio [14] and ANDSystem [15].

The ANDSystem was developed for the automated extraction of facts and knowledge regarding the relationships between proteins, genes, metabolites, microRNAs, cellular components, molecular processes, and their associations with diseases from published scientific texts and databases. To extract knowledge from texts in the ANDSystem, the shallow parsing method was applied. Pathway Studio is a software application developed for the navigation and analysis of biological pathways, gene regulation networks and protein interaction maps. The program uses the natural language processing approach to extract knowledge from the texts of scientific publications. STRING is a database and a web resource that contains information about protein-protein interactions (including physical and functional interactions) that is mainly based on the use of text-mining methods.

The associative networks for the considered disease pairs were graphs whose vertices were diseases and human proteins/genes, while the edges were the associations between diseases and proteins.

The following indices of relation between a pair of associative networks were used: (1) the intersection index, $I_{AB} = |A \cap B|$ equal to the intersection size of protein sets A and B composed of proteins concurrently associated with diseases D_A and D_B; (2) the Jaccard index [53] was calculated as the ratio of I_{AB} to the combination of sets A and B involving at least one of the diseases D_A and D_B, $J_{AB} = \frac{I_{AB}}{|A \cup B|}$; (3) Meet/Min [54] was calculated as $M_{AB} = \frac{I_{AB}}{\min(|A|, |B|)}$, where the denominator denotes the size of the minimum of sets A and B.

The statistical significance of the relation indices for the analysed diseases in the associative networks was determined by comparing these networks with the associative ones formed by pairs of randomly chosen diseases. For such an analysis, we used the ANDSystem because this program allows the comparison of reconstructed networks with random ones generated using the ANDCell knowledge base. All the interactions between proteins, genes, metabolites, diseases and other objects described by the ANDSystem are deposited in the ANDCell knowledge base, which is a module of this system [15]. The total number of diseases described in ANDCell was 4,075; of these, 991 were not found to be associated with any human protein. Such diseases were discarded from the analysis. To build the distribution of the relation indices for random disease pairs, 10,000 random disease pairs were generated (see Additional file 2). The P-value for the analysed disease pairs was calculated as the

proportion of 10,000 random networks with the same or larger CR as in the examined pairs of diseases. The associative networks were reconstructed using the ANDSystem and Pathway Studio programs. STRING was not used for this purpose because it gave no information regarding interactions between protein/gene and diseases. The associative networks for the analysed disease pairs included only interactions of the disease-protein/gene type; the interactions between proteins/genes were discounted. As a result, to analyse the interactions between proteins/genes in the associative networks, additional protein/gene-protein/gene associative molecular genetic networks were built using the ANDSystem, Pathway Studio and STRING. The statistical significance of the connectivity of the associative molecular genetics networks built for the analysed disease pairs was also determined by comparing them with random networks. In such a case, for each analysed associative molecular genetic networks, 1,000 random networks were generated using the ANDSystem (only human proteins/genes were considered).

The statistical significance (p-value) of the difference between the connectivity of the analysed network and that of the random networks was also determined, like in the case of the associative networks, as the proportion of random networks with the same or greater number of links between the vertices compared with the number of links in the analysed network. The random molecular genetic networks were built according to the following rules. Proteins/genes considered as vertices in the random networks were taken from the ANDCell knowledge base. To ensure that the proteins/genes in the random networks were represented at a level of study close to that of the proteins/genes from the analysed networks, we considered only those random proteins/genes whose connectivity rate was the same as connectivity rate of proteins/genes from the analysed networks. The set Q_i was formed for each i-th vertex of the analysed network. Q_i was composed of all the proteins/genes from the ANDCell knowledge base having an interaction number in ANDCell equal to the protein/gene interactions in the knowledge base represented by the i-th vertex. The protein/gene for the i-th vertex of the random network was chosen by chance for the set Q_i . The links between the vertices in the random networks were set according to the interactions described in the ANDCell knowledge base.

The results of the automated extraction of information regarding the interactions between proteins/genes and diseases were tested manually. The recognition correctness of the object names in the text, as well as the presence of their interactions, was tested. The lists of shared and specific proteins were reduced by expert evaluation to retain only those participating in the pathogenesis of both diseases for shared proteins, and in the pathogenesis of either disease for specific proteins, as shown previously [10].

The BINGO tool [55] was used to evaluate the overrepresentation of the biological processes for the considered protein/gene set. The enrichment was evaluated using a hypergeometric test with the Benjamini and Hochberg FDR correction using the whole annotation as a reference set. The human Uniprot-GOA Gene Association file (release 2013_05) was used as the custom annotation file. In addition to the statistical significance of the overrepresentation, the overrepresented GO processes were characterized by the CR of the respective proteins/genes in the associative molecular genetics network built for intersection of the four studied diseases. The CR for the protein group of the examined network involved in the overrepresented GO biological process was calculated as the ratio of the number of the protein pairs connected by the network protein pairs of the given group to the number of all possible pairwise combinations of proteins of this group. As is known, the reconstruction quality of the molecular genetic networks is related frequently to the problem of the completeness of information regarding the interactions between proteins. For this reason, to build the network, we took advantage of three independent programs: ANDSystem, Pathway Studio and STRING, with their parameters set by default.

Additional material

Additional file 1: Excel spreadsheet file containing information regarding the characteristics of overrepresented biological processes.

Additional file 2: Excel spreadsheet file containing information regarding the distribution of the relation indices of the disease-protein-disease associative networks.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Expert analysis of the pathogenetic contributors (diabetes mellitus, gestational diabetes and obesity) was done by ASG, ESV, VSP, ONA, EVM, MSZ and VSB. The development of methods, programs, calculations and analyses of the structural organization of the molecular genetic networks was done by EST, PSD, OVS, TVI, NAK and VAI. All authors read and approved the final manuscript.

Acknowledgements

The work was supported in part by the Russian Science Foundation grant No. 14-24-00123 (development of methods, programs and reconstruction and analysis of the pre-eclampsia associative networks) and Saint-Petersburg State University grant No. 1.38.79.2012 (expert analysis of the pathogenetic contributors: diabetes mellitus, gestational diabetes and obesity).

Declarations

Publication of this article has been funded by the Russian Science Foundation grant No. 14-24-00123.

This article has been published as part of *BMC Systems Biology* Volume 9 Supplement 2, 2015: Selected articles from the IX International Conference on the Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS \SB-2014): Systems Biology. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcsystbiol/supplements/9/S2.

Authors' details
[1]Federal State Budget scientific Institution "#34;The Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott", St. Petersburg, Russia. [2]Saint-Petersburg State University, St. Petersburg, Russia. [3]The Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia. [4]Novosibirsk State University, Novosibirsk, Russia.

Published: 15 April 2015

References

1. Bilano VL, Ota E, Ganchimeg T, Mori R, Souza JP: **Risk factors of pre-eclampsia/eclampsia and its adverse outcomes in low- and middle-income countries: a WHO secondary analysis.** *PLoS One* 2014, **9**:e91198.
2. Chaiworapongsa T, Chaemsaithong P, Yeo L, Romero R: **Pre-eclampsia part 1: current understanding of its pathophysiology.** *Nat Rev Nephrol* 2014.
3. Young BC, Levine RJ, Karumanchi SA: **Pathogenesis of pre-eclampsia.** *Annu Rev Pathol* 2010, **5**:173-92.
4. Duckitt K, Harrington D: **Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies.** *BMJ* 2005, **330**:565.
5. Bellamy L, Casas J-PP, Hingorani AD, Williams DJ: **Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis.** *BMJ* 2007, **335**:974.
6. Feinstein AR: **The pre-therapeutic classification of co-morbidity in chronic disease.** *Journal of Chronic Diseases* 1970, **23**:455-468.
7. Pfaundler M, Seht L: **Über Syntropie von Krankheitszuständen.** *Zeitschrift für Kinderheilkunde* 1921, **30**:100-120.
8. Ibáñez K, Boullosa C, Tabarés-Seisdedos R, Baudot A, Valencia A: **Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses.** *PLoS genetics* 2014, **10**:e1004173.
9. Freidin MB, Puzyrev VP: *Syntropic genes of allergic diseases* 2010.
10. Bragina EY, Tiys ES, Freidin MB, Koneva LA, Demenkov PS, Ivanisenko VA, Kolchanov NA, Puzyrev VP: **Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis.** *Immunogenetics* 2014, **66**:457-65.
11. Jenssen TK, Laegreid A, Komorowski J, Hovig E: **A literature network of human genes for high-throughput analysis of gene expression.** *Nat Genet* 2001, **28**:21-8.
12. Cheung WA, Ouellette BFF, Wasserman WW: **Quantitative biomedical annotation using medical subject heading over-representation profiles (MeSHOPs).** *BMC bioinformatics* 2012, **13**:249.
13. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Mering C von, Jensen LJ: **STRING v9.1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic Acids Res* 2013, **41(Database)**:D808-15.
14. Nikitin A, Egorov S, Daraselia N, Mazo I: **Pathway studio--the analysis and navigation of molecular networks.** *Bioinformatics* 2003, **19**:2155-7.
15. Demenkov PS, Ivanisenko TV, Kolchanov NA, Ivanisenko VA: **ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem.** *In Silico Biol* 2011, **11**:149-61.
16. Larina IM, Kolchanov NA, Dobrokhotov IV, Ivanisenko VA, Demenkov PS, Tiĭs ES, Valeeva OA, Pastushkova LK, Nikolaev EN: **[Reconstruction of associative protein networks connected with processes of sodium exchange' regulation and sodium deposition in healthy volunteers by urine proteome analysis].** *Fiziol Cheloveka* 2012, **38**:107-15.
17. Momynaliev KT, Kashin SV, Chelysheva VV, Selezneva OV, Demina IA, Serebryakova MV, Alexeev D, Ivanisenko VA, Aman E, Govorun VM: **Functional divergence of Helicobacter pylori related to early gastric cancer.** *Journal of proteome research* 2010, **9**:254-67.
18. Podkolodnaya OA, Yarkova EE, Demenkov PS: **Application of the ANDCell computer system to reconstruction and analysis of associative networks describing potential relationships between myopia and glaucoma.** *Russian Journal of Genetics: Applied Research* 2011, **1**:21-28.
19. Sommer B, Tiys ES, Kormeier B, Hippe K, Janowski SJ, Ivanisenko TV, Bragin AO, Arrigo P, Demenkov PS, Kochetov AV, Ivanisenko VA, Kolchanov NA, Hofestädt R: **Visualization and analysis of a cardio vascular disease- and MUPP1-related biological network combining text mining and data warehouse approaches.** *J Integr Bioinform* 2010, **7**:148.
20. Puzyrev VP, Freidin MB: **Genetic view on the phenomenon of combined diseases in man.** *Acta Naturae* 2009, **1**:52-7.
21. Mahaba HM, Ismail NA, El Damaty SI, Kamel HA: **Pre-eclampsia: epidemiology and outcome of 995 cases.** *J Egypt Public Health Assoc* 2001, **76**:357-68.
22. Wendland EM, Duncan BB, Belizán JM, Vigo A, Schmidt MI: **Gestational diabetes and pre-eclampsia: common antecedents?** *Arq Bras Endocrinol Metabol* 2008, **52**:975-84.
23. Schneider S, Freerksen N, Röhrig S, Hoeft B, Maul H: **Gestational diabetes and pre-eclampsia–similar risk factor profiles?** *Early Hum Dev* 2012, **88**:179-84.
24. Ornaghi S, Tyurmorezova A, Algeri P, Giardini V, Ceruti P, Vertemati E, Vergani P: **Influencing factors for late-onset pre-eclampsia.** *J Matern Fetal Neonatal Med* 2013, **26**:1299-302.
25. Lisonkova S, Joseph KS: **Incidence of pre-eclampsia: risk factors and outcomes associated with early-versus late-onset disease.** *Am J Obstet Gynecol* 2013, **209**:544-e1.
26. Dadelszen P von, Magee LA: **Pre-eclampsia: an update.** *Curr Hypertens Rep* 2014, **16**:454.
27. Lučovnik M, Tul N, Verdenik I, Novak Z, Blickstein I: **Risk factors for pre-eclampsia in twin pregnancies: a population-based matched case-control study.** *J Perinat Med* 2012, **40**:379-82.
28. Buurma AJ, Turner RJ, Driessen JH, Mooyaart AL, Schoones JW, Bruijn JA, Bloemenkamp KW, Dekkers OM, Baelde HJ: **Genetic variants in pre-eclampsia: a meta-analysis.** *Hum Reprod Update* 2013, **19**:289-303.
29. Dostálová Z, Bienertová-Vasků AJ, Vasků A, Gerychová R, Unzeitig V: **[Insertion-deletion polymorphism in the gene for angiotensin-converning enzyme (I/D ACE) in pregnant women with gestational diabetes].** *Ceska Gynekol* 2006, **71**:369-73.
30. Yu Z-YY, Chen L-SS, Zhang L-CC, Zhou T-BB: **Meta-analysis of the relationship between ACE I/D gene polymorphism and end-stage renal disease in patients with diabetic nephropathy.** *Nephrology (Carlton)* 2012, **17**:480-7.
31. Mao S, Huang S: **A meta-analysis of the association between angiotensin-converting enzyme insertion/ deletion gene polymorphism and the risk of overweight/obesity.** *J Renin Angiotensin Aldosterone Syst* 2013.
32. Machado JS, Palei AC, Amaral LM, Bueno AC, Antonini SR, Duarte G, Tanus-Santos JE, Sandrim VC, Cavalli RC: **Polymorphisms of the adiponectin gene in gestational hypertension and pre-eclampsia.** *J Hum Hypertens* 2014, **28**:128-32.
33. Low CF, Mohd Tohit ER, Chong PP, Idris F: **Adiponectin SNP45TG is associated with gestational diabetes mellitus.** *Arch Gynecol Obstet* 2011, **283**:1255-60.
34. Chu H, Wang M, Zhong D, Shi D, Ma L, Tong N, Zhang Z: **AdipoQ polymorphisms are associated with type 2 diabetes mellitus: a meta-analysis study.** *Diabetes Metab Res Rev* 2013, **29**:532-45.
35. Wu J, Liu Z, Meng K, Zhang L: **Association of adiponectin gene (ADIPOQ) rs2241766 polymorphism with obesity in adults: a meta-analysis.** *PloS One* 2014, **9**:e95270.
36. Vianna P, Silva GK Da, Santos BP Dos, Bauer ME, Dalmáz CA, Bandinelli E, Chies JA: **Association between mannose-binding lectin gene polymorphisms and pre-eclampsia in Brazilian women.** *Am J Reprod Immunol (New York, NY: 1989)* 2010, **64**:359-74.
37. Megia A, Gallart L, Fernández-Real J-MM, Vendrell J, Simón I, Gutierrez C, Richart C: **Mannose-binding lectin gene polymorphisms are associated with gestational diabetes mellitus.** *J Clin Endocrinol Metab* 2004, **89**:5081-7.
38. Muller YL, Hanson RL, Bian L, Mack J, Shi X, Pakyz R, Shuldiner AR, Knowler WC, Bogardus C, Baier LJ: **Functional variants in MBL2 are associated with type 2 diabetes and pre-diabetes traits in Pima Indians and the old order Amish.** *Diabetes* 2010, **59**:2080-5.
39. Fernández-Real JM, Straczkowski M, Vendrell J, Soriguer F, Pérez Del Pulgar S, Gallart L, López-Bermejo A, Kowalska I, Manco M, Cardona F, García-Gil MM, Mingrone G, Richart C, Ricart W, Zorzano A: **Protection from inflammatory disease in insulin resistance: the role of mannan-binding lectin.** *Diabetologia* 2006, **49**:2402-11.

40. Harmon QE, Engel SM, Wu MC, Moran TM, Luo J, Stuebe AM, Avery CL, Olshan AF: **Polymorphisms in inflammatory genes are associated with term small for gestational age and pre-eclampsia.** *Am J Reprod Immunol* 2014, **71**:472-84.
41. Chang Y, Niu XM, Qi XM, Zhang HY, Li NJ, Luo Y: **[Study on the association between gestational diabetes mellitus and tumor necrosis factor-alpha gene polymorphism].** *Zhonghua Fu Chan Ke Za Zhi* 2005, **40**:676-8.
42. Sefri H, Benrahma H, Charoute H, Lakbakbi El Yaagoubi F, Rouba H, Lyoussi B, Nourlil J, Abidi O, Barakat A: **TNF A -308G>A polymorphism in Moroccan patients with type 2 diabetes mellitus: a case-control study and meta-analysis.** *Mol Biol Rep* 2014.
43. Sookoian SC, González C, Pirola CJ: **Meta-analysis on the G-308A tumor necrosis factor alpha gene variant and phenotypes associated with the metabolic syndrome.** *Obes Res* 2005, **13**:2122-31.
44. Ehsanipoor RM, Fortson W, Fitzmaurice LE, Liao W-XX, Wing DA, Chen D-BB, Chan K: **Nitric oxide and carbon monoxide production and metabolism in pre-eclampsia.** *Reprod Sci* 2013, **20**:542-8.
45. Lachmeijer AM, Nosti-Escanilla MP, Bastiaans EB, Pals G, Sandkuijl LA, Kostense PJ, Aarnoudse JG, Crusius JB, Peña AS, Dekker GA, Arngrímsson R, Kate LP ten: **Linkage and association studies of IL1B and IL1RN gene polymorphisms in pre-eclampsia.** *Hypertens Pregnancy* 2002, **21**:23-38.
46. Serrano NC, Casas JP, Díaz LA, Páez C, Mesa CM, Cifuentes R, Monterrosa A, Bautista A, Hawe E, Hingorani AD, Vallance P, López-Jaramillo P: **Endothelial NO synthase genotype and risk of pre-eclampsia: a multicenter case-control study.** *Hypertension* 2004, **44**:702-7.
47. Chen M, Yuan Z, Shan K: **Association of apolipoprotein J gene 866C–>T polymorphism with pre-eclampsia and essential hypertension.** *Gynecol Obstet Invest* 2005, **60**:133-8.
48. Fekete A, Vér A, Bögi K, Treszl A, Rigó J: **Is pre-eclampsia associated with higher frequency of HSP70 gene polymorphisms?** *Eur J Obstet Gynecol Reprod* 2006, **126**:197-200.
49. Mütze S, Rudnik-Schöneborn S, Zerres K, Rath W: **Genes and the pre-eclampsia syndrome.** *J Perinat Med* 2008, **36**:38-58.
50. Wang XM, Wu HY, Qiu XJ: **Methylenetetrahydrofolate reductase (MTHFR) gene C677T polymorphism and risk of pre-eclampsia: an updated meta-analysis based on 51 studies.** *Arch Med Res* 2013, **44**:159-68.
51. Vanderweele TJ: **Sensitivity analysis: distributional assumptions and confounding assumptions.** *Biometrics* 2008, **64**:645-9.
52. Vanderweele TJ, Mukherjee B, Chen J: **Sensitivity analysis for interactions under unmeasured confounding.** *Stat Med* 2012, **31**:2552-64.
53. Jaccard P: **The distribution of the flora in the alpine zone.** *New Phytol* 1912, **11**:37-50.
54. Goldberg DS, Roth FP: **Assessing experimentally derived interactions in a small world.** *Proc Natl Acad Sci USA* 2003, **100**:4372-4376.
55. Maere S, Heymans K, Kuiper M: **BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.** *Bioinformatics* 2005, **21**:3448-9.

doi:10.1186/1752-0509-9-S2-S4
Cite this article as: Glotov *et al.*: Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome). *BMC Systems Biology* 2015 **9**(Suppl 2):S4.

OPEN ACCESS Freely available online

PLOS ONE

Toxicological Effects of the Different Substances in Tobacco Smoke on Human Embryonic Development by a Systems Chemo-Biology Approach

Bruno César Feltes[1], Joice de Faria Poloni[2], Daniel Luis Notari[3], Diego Bonatto[1]*

1 Department of Molecular Biology and Biotechnology, Biotechnology Center of the Federal University of Rio Grande do Sul, Federal University of Rio Grande do Sul, Porto Alegre, RS – Brazil, **2** Institute of Biotechnology, University of Caxias do Sul, Caxias do Sul, RS – Brazil, **3** Computational and Information Technology Center, Universidade de Caxias do Sul, Caxias do Sul, RS – Brazil

Abstract

The physiological and molecular effects of tobacco smoke in adult humans and the development of cancer have been well described. In contrast, how tobacco smoke affects embryonic development remains poorly understood. Morphological studies of the fetuses of smoking pregnant women have shown various physical deformities induced by constant fetal exposure to tobacco components, especially nicotine. In addition, nicotine exposure decreases fetal body weight and bone/cartilage growth in addition to decreasing cranial diameter and tibia length. Unfortunately, the molecular pathways leading to these morphological anomalies are not completely understood. In this study, we applied interactome data mining tools and small compound interaction networks to elucidate possible molecular pathways associated with the effects of tobacco smoke components during embryonic development in pregnant female smokers. Our analysis showed a relationship between nicotine and 50 additional harmful substances involved in a variety of biological process that can cause abnormal proliferation, impaired cell differentiation, and increased oxidative stress. We also describe how nicotine can negatively affect retinoic acid signaling and cell differentiation through inhibition of retinoic acid receptors. In addition, nicotine causes a stress reaction and/or a pro-inflammatory response that inhibits the agonistic action of retinoic acid. Moreover, we show that the effect of cigarette smoke on the developing fetus could represent systemic and aggressive impacts in the short term, causing malformations during certain stages of development. Our work provides the first approach describing how different tobacco constituents affect a broad range of biological process in human embryonic development.

Citation: Feltes BC, Poloni JdF, Notari DL, Bonatto D (2013) Toxicological Effects of the Different Substances in Tobacco Smoke on Human Embryonic Development by a Systems Chemo-Biology Approach. PLoS ONE 8(4): e61743. doi:10.1371/journal.pone.0061743

Editor: Michael Schubert, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, France

Received October 12, 2012; **Accepted** March 15, 2013; **Published** April 29, 2013

Funding: This work was supported by research grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Grant Number 474117/2010-3), the Programa Institutos Nacionais de Ciência e Tecnologia (INCT de Processos Redox em Biomedicina-REDOXOMA; Grant Number 573530/2008-4; http://www.cnpq.br), Fundação de Amparo a Pesquisa do Rio Grande do Sul FAPERGS (PRONEM Grant Number 11/2072-2; http://www.fapergs.rs.gov.br) and CAPES (Cordenação de Aperfeiçoamento de Pessoal do Ensino Superior; http://www.capes.gov.br). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: diegobonatto@gmail.com

Introduction

There are more than 4,800 compounds present in the particulate and vapor phases of cigarette smoke [1], and many of these compounds are considered to represent a human health risk [2]. Known constituents of cigarette smoke include isoprene, butadiene, polycyclic aromatic hydrocarbons (PAHs), aldehydes, metals, *N*-nitrosamines, and aromatic amines, in addition to many others [1]. Although extensive anti-tobacco public advertisements promote smoking cessation in pregnant women, a considerable number of women still smoke during their pregnancies and/or are exposed to tobacco smoke via passive smoking [2], [3], [4].

We addressed two major issues in this work. Although prenatal smoke exposure has been previously associated with innumerable malformations during fetus growth and development and disruptions of reproductive physiology, there are gaps in the knowledge of how tobacco components (TCs) affect the developing embryo in pregnant women in a systemic way, [2], [3], [5], [6]. This knowledge gap is the first issue that we address. Interestingly, these abnormalities are not tissue specific or related to any unique pathway but, rather, are systemic and connected to a broad range of birth defects [2], [4]. The second issue that we address relates to the fact that nicotine is the principal psychoactive constituent of tobacco, understanding its biological effects on fetal and maternal health is critical, as it may affect distinct biochemical pathways when compared to other tobacco smoke constituents. Studies concerning the morphological effects of tobacco smoke constituents in fetuses from both active and passive smoking women have shown significant alterations in weight, fat mass and most anthropometric parameters as well as in the placenta with alterations in protein metabolism and enzyme activity [7]. These alterations are the results of a direct toxic effect on the fetal cells or an indirect effect through damage to, and/or functional disturbances of the placenta [7]. One possible explanation that could link nicotine and the negative regulation of development is retinoic acid (RA) signaling. RA is an indispensable molecule involved in the regulation of gene expression and cell-cell signaling during early development [8]. RA can cross the cell membrane and bind

to specific nuclear receptors, such as retinoic acid receptors (RARs) and retinoid × receptors (RXRs) [8]. Studies regarding the role of RA receptors during embryogenesis have shown that RARs are essential for the expression of HOX genes and skeletal development [8], [9]. Nicotine has been previously associated with inhibition of the RARβ gene in lung cancer, which suggests that nicotine affects RA signaling in human tissues [10]. Therefore, RA signaling is a plausible pathway through which nicotine could affect cell differentiation and cause human fetal morphological abnormalities. However, the molecular mechanisms underlying the progression or the cause of fetal abnormalities related to cigarette smoking remain unknown.

To understand these mechanisms, we performed systems chemo-biology analyses to elucidate the nature and number of proteins and modules that are associated with prenatal tobacco smoke exposure. Different protein-protein interaction (PPI) and chemical-protein interaction (CPI) networks derived from interactome projects were described. In a first analysis, we prospected and analyzed a network using a list of 95 commonly found harmful tobacco constituents [2], to elucidate how these substances could act together to influence embryonic and fetal development. In a second systems chemo-biology analysis, we prospected data on the interactome and small compounds for nicotine alone and examined how they could negatively affect cell differentiation and bone development and lead to morphological abnormalities. Furthermore, we conducted gene ontology (GO) analyses of the major biological processes derived from the PPI and CPI networks. Supporting the hypotheses gathered from systems chemo-biology analyses, a landscape network study was performed using available transcriptomic data of placenta and cord blood isolated from passive smoking women and non-smoking women [11].

A model of how selected TCs could influence embryonic development was generated. We also developed a separate model of how nicotine could affect cell differentiation and bone development. Taken together, our systems chemo-biology data are the first to show how tobacco smoke can affect fetal and embryonic development in a systemic matter at the molecular level.

Materials and Methods

Interactome Data Mining and Design of the Chemo-biology Network

To design chemo-biology interactome networks and to elucidate the interplay between development and TCs, the metasearch engines STITCH 3.1 [http://stitch.embl.de/] and STRING 9.0 [http://string-db.org/] [12], [13] were used. In this sense, a list of 51 commonly found TCs, many of them with known concentrations in the mainstream and sidestream tobacco smoke [2] were used as initial seed for network prospection in STITCH. STITCH software allows visualization of the physical connections among different proteins and chemical compounds, whereas STRING shows protein-protein interactions. Each protein-protein or protein-chemical connection (edge) shows a degree of confidence between 0 and 1.0 (with 1.0 indicating the highest confidence). The parameters used in STITCH software were as follows: all prediction methods enabled, excluding text mining; 20 to 50 interactions; degree of confidence, medium (0.400); and a network depth equal to 1. The results gathered using these search engines were analyzed with Cytoscape 2.8.2 [14]. In addition, the GeneCards [http://www.genecards.org/] [15], [16], KEGG [http://www.genome.jp/kegg/] [17], iHop [http://www.ihop-net.org/UniPub/iHOP/] [18], PubChem [http://pubchem.ncbi.nlm.nih.gov/], ALOGPS 2.1 [http://www.vcclab.org/lab/alogps/] [19], AmiGO 1.8 [http://amigo.geneontology.org/cgi-bin/amigo/go.cgi] [20], and Gene Expression Atlas [http://www.ebi.ac.uk/gxa/] [21] search engines were also employed using their default parameters.

To prospect protein-protein and chemical-protein interactions (PPI and CPI, respectively), we entered each TC into the STITCH program. TCs that were not present in the STITCH database (or those that did not shown any protein connections) and particularly well described components, such as nitric oxide, phenol and carbon monoxide, were excluded from the analysis.

Different small CPI and PPI networks were obtained (data not shown), and these networks were further analyzed using Cytoscape 2.8.2. Each network generated by STITCH and STRING was combined into a large network using the Advanced Merge Network function, which was fully implemented in Cytoscape software.

Gene Expression Data for the Main Associated Nodes of Tobacco Components

To determine whether mRNA sequences associated with specific proteins connected to each TC could be present during development, we searched the transcriptome data from the Gene Expression Atlas [22]. We used the protein name and expression data for *Homo sapiens* embryos and fetuses as the initial inputs. The expression data indicated overexpressed and underexpressed genes (Table S1 in Supporting Information S1). Gene Expression Atlas infers the expression data for a specified gene by providing a list of experimental studies [22]. We considered a gene overexpressed or underexpressed based on the number of studies that matched the expression state of our input. Proteins that are only present in embryonic tissue were colored green, whereas proteins that are only present in fetal tissue were colored pink (Table S1 in Supporting Information S1). The blue nodes indicate the presence of a protein in both embryonic and fetal tissue (Table S1 in Supporting Information S1). Uncolored nodes (default color white) connected to TCs were either not present in any of the selected tissues in the initial input or were not found in the Gene Expression Atlas database (**Fig. 1**).

Additionally, we evaluated the transcriptomic data gathered from placenta and cord blood of passive smoking women (termed group "a"), with cord blood cotinine levels >1.0 ng/mL, and from non-smoking women (group "b"), with cord blood cotinine levels <0.15 ng/mL [11]. For this purpose, the matrix file GSE30032 (available at Gene Expression Omnibus [http://www.ncbi.nlm.nih.gov/geo]) was used and a mean value of expression for each gene was generated for both groups "a" and "b". The mean value of expression was then overlaid in CPI-PPI-derived subnetworks with the software ViaComplex 1.0 [23]. By providing gene expression data and interactomic networks, the software ViaComplex generates a landscape view of gene expression in a specific network.

Solubility Predictions for Major Tobacco Component-associated CPI-PPI Networks

To predict the solubility of each TC in an aqueous environment, such as in blood and plasma, we used the program ALOGPS 2.1 [http://www.vcclab.org/lab/alogps/]. ALOGPS allows simulation of the probable solubility of a given compound determined based on its structural formula or CAS number. Compounds with a solubility of less than 35 g/L [values of ALOGpS and logS (exp)] were considered lipophilic. ALOGPS 2.1 was used with its default parameters.

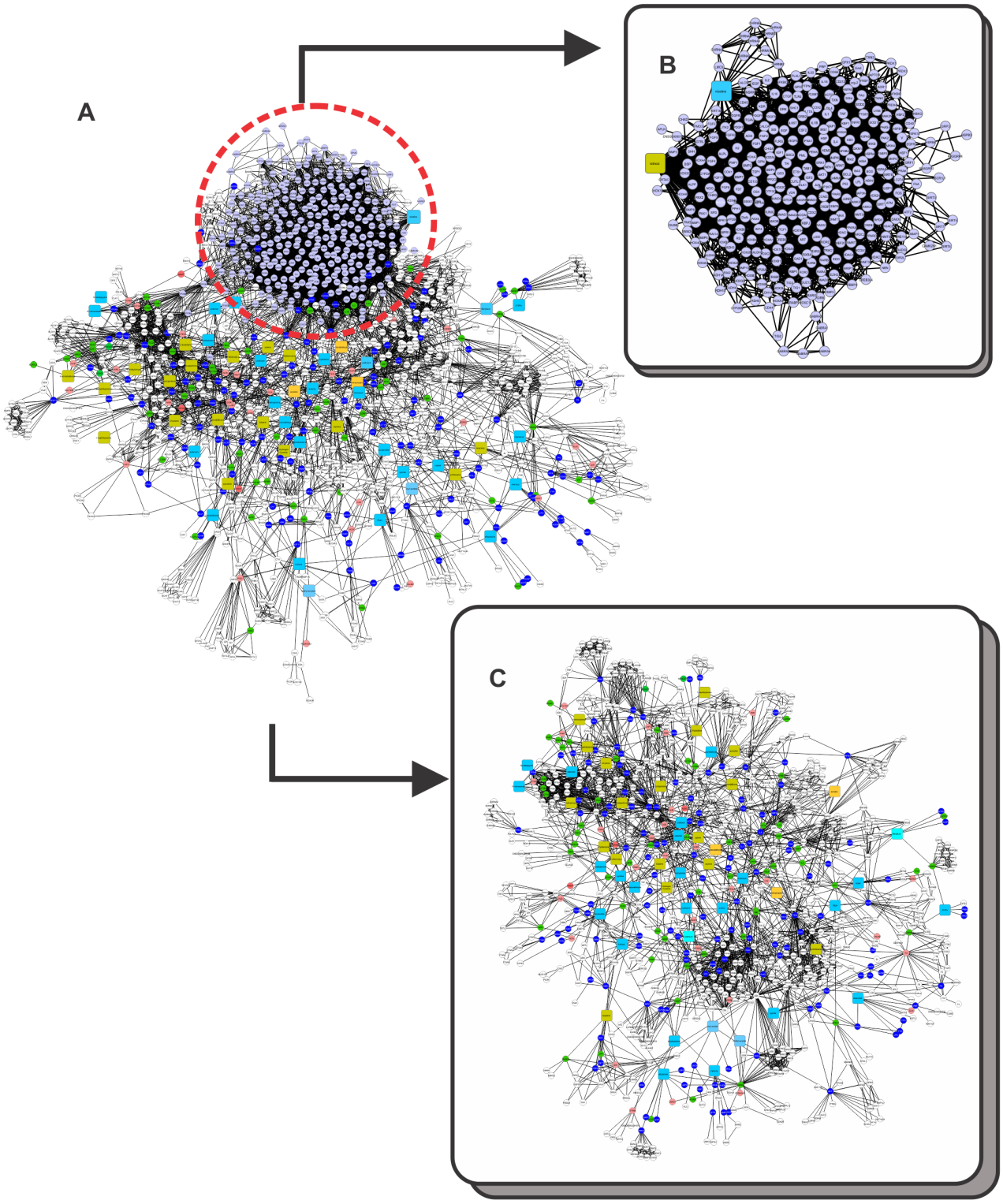

Figure 1. A binary network of chemical-protein and protein-protein interactions (CPI-PPI network) generated by the program Cytoscape 2.8.2. (A) The main network, showing 49 known substances present in tobacco, 1177 nodes (49 substances, 1128 proteins) and 7522 edges (connections). Proteins were colored to identify the tissue in which they were present: (i) pink indicates fetal tissue; (ii) green, embryonic tissue; and (iii) dark blue, both fetal and embryonic tissues. In addition, each substance was colored according to its solubility: (i) yellow indicates lipophilic and (ii) light blue, hydrophilic. We observed that nicotine resided in a module apart from the major network (A). Therefore, we separated it from the major CPI-PPI network and colored its module purple. (B) The nicotine subnetwork is shown separately from the major CPI-PPI network. It contained proteins related to retinoic acid signaling and retinoic acid (lipophilic molecule). (C) The final major CPI-PPI network after the nicotine module was extracted.

doi:10.1371/journal.pone.0061743.g001

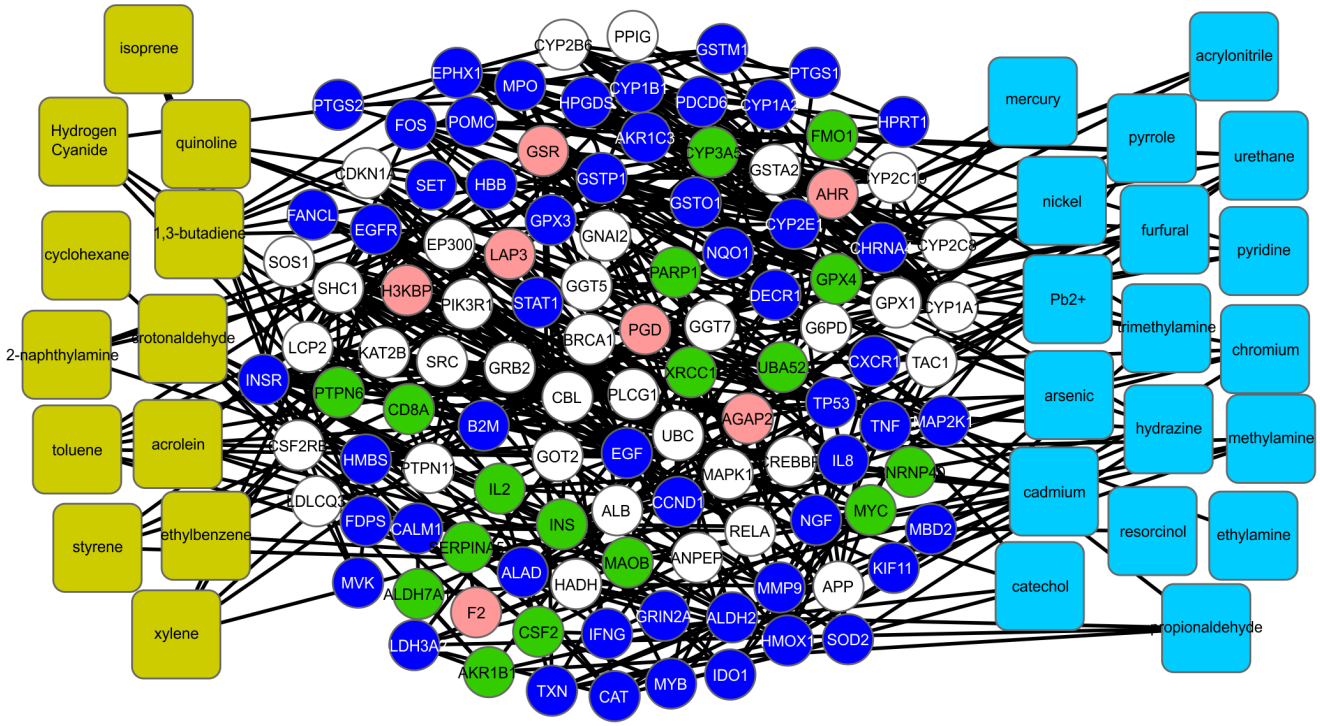

Figure 2. HBs found in the major CPI-PPI network. Betweenness and node degrees were assessed using the program CentiScaPe. Among the 143 HBs, 53 proteins are present in both tissues, reinforcing the idea of a prolonged effect of CS on embryonic development.
doi:10.1371/journal.pone.0061743.g002

Module Analysis of Major Tobacco Component-associated CPI-PPI Networks

The large CPI-PPI network obtained from the initial search (**Fig. 1**) was analyzed in terms of the major cluster or module composition using the program Molecular Complex Detection (MCODE) [24], which is available at http://baderlab.org/Software/MCODE. MCODE is based on vertex weighting by the local neighborhood density and outward traversal from a locally dense seed protein to isolate the dense regions according to given parameters stipulated by the researcher [24]. The parameters for cluster finding were as follows: loops included; degree cutoff, 2; expansion of a cluster by one neighbor shell allowed (fluff option enabled); deletion of a single connected node from clusters (haircut option enabled); node density cutoff, 0.1; node score cutoff, 0.2; kcore, 2; and maximum network depth, 100. Each cluster generates a value of "cliquishness" (C*i*), which is the degree of connection in a given group of proteins. Thus, the higher the C*i* value, the more connected the cluster [24].

Centrality Analysis of the Major Tobacco Component-associated CPI-PPI Networks

Centrality analysis was performed using the program CentiScaPe 1.2 [25]. In this analysis, the CentiScaPe algorithm evaluates each network node according to the node degree, betweenness and closeness to establish the most "central" nodes (proteins/chemicals) within the network. Thus, the most relevant node for a determined biochemical pathway or module can be obtained and further analyzed. In general terms, the closeness analysis (1) indicates the probability that any protein/chemical compound (node in our network) is relevant to another protein/chemical compound (node) in a signaling network or its associated network [25], as determined using Equation (1):

$$Clo(v) = \frac{1}{\sum_{w \in V} dist(v,w)} \quad (1)$$

where the closeness value of node v ($Clo(v)$) is determined by computing and totalizing the shortest paths among node v and all other nodes (w; $dist(v,w)$) found within a network (1). The average closeness (Clo) score was obtained by calculating the sum of different closeness scores (Clo_i) divided by the total number of nodes analyzed ($N(v)$) (Equation 2).

$$\langle Clo \rangle = \frac{\sum_i Clo_i}{N_{(v)}} \quad (2)$$

The higher the closeness value compared to the average closeness score, the higher the relevance of the protein/chemical compound to other protein nodes within the network/module. In turn, the betweenness indicates the number of the shortest paths that go through each node (Equation 3) [25], [26]:

$$Bet(v) = \sum_{s \neq v \neq t} \frac{\sigma_{sw}(v)}{\sigma_{sw}} \quad (3)$$

where σ_{stw} total number of the shortest paths from node s to node w, and σ_{sw} (v) is the number of those paths that pass through the node. The average betweenness score (Bet) of the network was calculated using equation (4), where the sum of different betweenness scores (Bet_i) is divided by the total number of nodes analyzed ($N(v)$):].

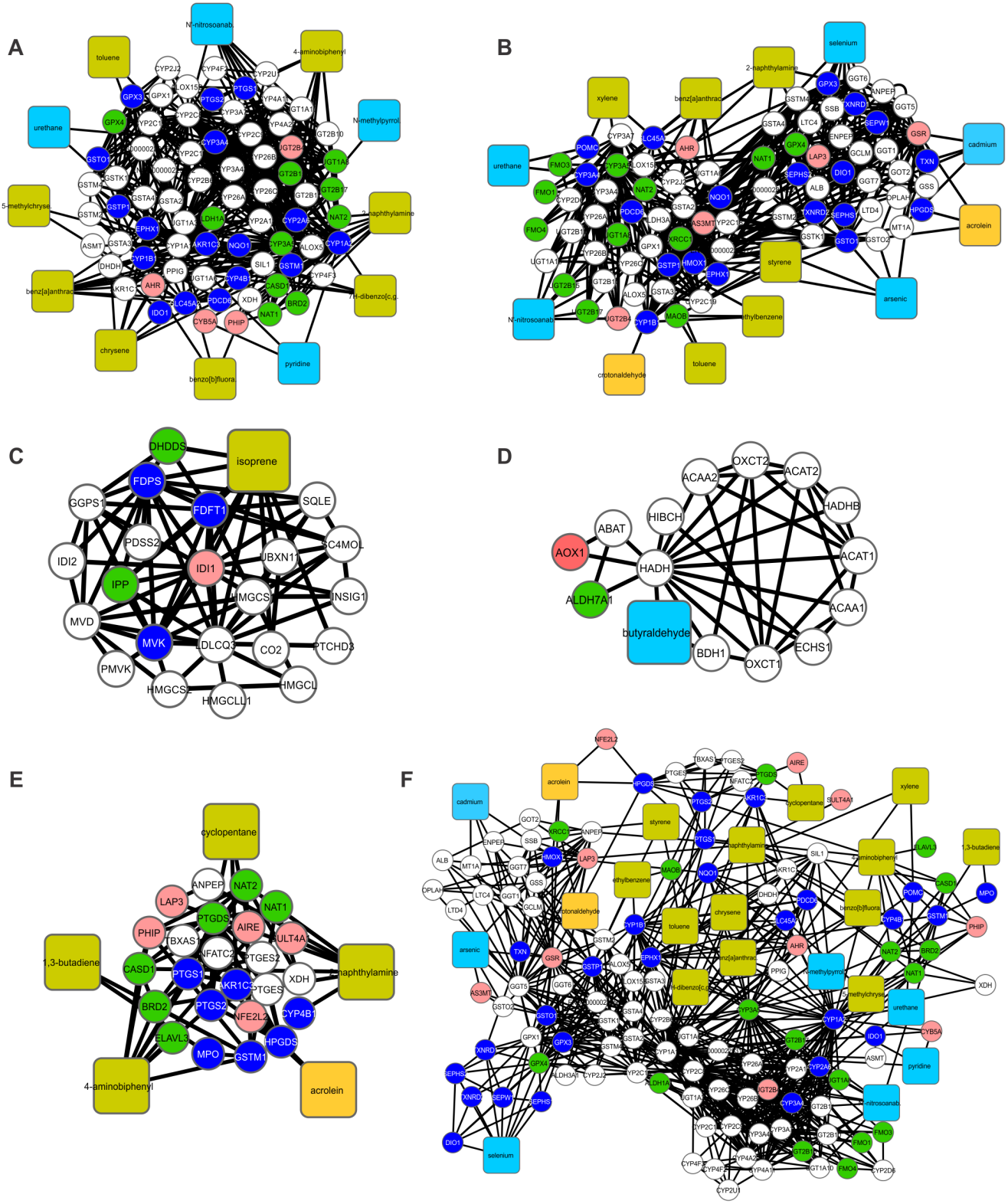

Figure 3. Cluster analysis of the major CPI-PPI network indicating clusters 1, 4, 11, 16 and 20. Cluster 1 (A) is composed of 83 nodes and 565 edges, with *Ci* = 6,843. The associated hydrophilic constituents are urethane, *N*-nitrosoanabasine, *N*-methylpyrrolidine and pyridine. The lipophilic constituents are toluene, 4-aminobiphenyl, 5-methylcrysene, benz[a]anthracene, chrysene, benzo[b]fluoranthene, 7H-dibenzo[c,g]carbazole and 2-naphthylamine. Related GO terms: oxidation reduction and unsaturated fatty acid metabolic processes. Cluster 4 (B) is composed of 90 nodes and 411 edges, with *Ci* = 4,567. The associated hydrophilic compounds are urethane, *N*-nitrosoanabasine, *N*-methylpyrrolidine, arsenic, selenium and cadmium, and the lipophilic compounds are acrolein, crotonaldehyde, toluene, xylene, ethylbenzene, benz[a]anthracene, styrene and 2-naphthylamine. Related GO term: oxidation reduction. Cluster 11 (C) is composed of 23 nodes and 74 edges, with *Ci* = 3,217. Only the lipophilic

compound isoprene is present in this cluster. Related GO term: steroid biosynthetic processes. Cluster 16 (D) is composed of 15 nodes and 36 edges, with $Ci = 2{,}400$. The associated hydrophilic compound is butyraldehyde. Related GO term: lipid modification. Cluster 20 (E) is composed of 29 nodes and 65 edges, with $Ci = 2{,}241$. The associated lipophilic compounds are acrolein, 2-naphthylamine, 1,3-butadiene, cyclopentane and 4-aminobiphenyl. Related GO terms: prostaglandin metabolic processes and unsaturated fatty acid metabolic processes. A merge of clusters 1, 4 and 20 (F). Clusters 11 and 16 did not show any proteins overlapping with any other cluster.
doi:10.1371/journal.pone.0061743.g003

$$\langle Bet \rangle = \frac{\sum_i Bet_i}{N_{(v)}} \quad (4)$$

Thus, nodes with high betweenness scores compared to the average betweenness score of the network are responsible for controlling the flow of information through the network topology. The higher a node's betweenness score, the higher the probability that the node connects different modules or biological processes, such nosed are called bottleneck nodes.

Finally, the node degree ($Deg(v)$) is a measure that indicates the number of connections (Ei) that involve a specific node (v) (Equation 5):

$$Deg(v) = \sum E_i \quad (5)$$

The average node degree of a network (Deg) is given by equation 6, where the sum of different node degree scores (Bet_i) is divided by the total number of nodes ($N(v)$) present in the network:

$$\langle Deg \rangle = \frac{\sum_i Deg_i}{N_{(v)}} \quad (6)$$

Nodes with a high node degree are called hubs [25] and have key regulatory functions in the cell.

Gene Ontology Analyses of Major Tobacco Component-associated CPI-PPI Networks

The CPI-PPI modules generated by MCODE were further studied by focusing on major biology-associated processes using the Biological Network Gene Ontology (BiNGO) 2.44 Cytoscape plugin [27], available at http://www.cytoscape.org/plugins2.php#IO_PLUGINS. The degree of functional enrichment for a given cluster and category was quantitatively assessed (p-value) using a hypergeometric distribution. Multiple test correction was also assessed by applying the false discovery rate (FDR) algorithm [28], which was fully implemented in BiNGO software at a significance level of $p<0.05$. The most statistically relevant processes were taken into account when developing the interaction model.

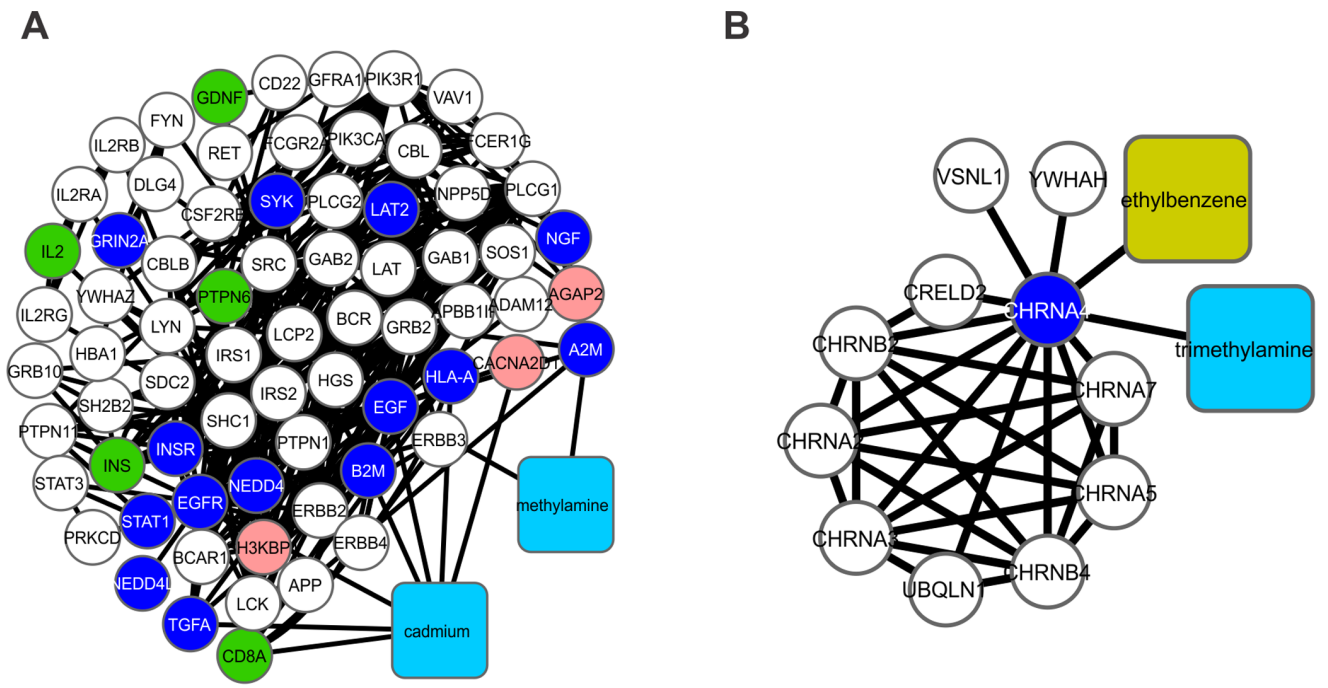

Figure 4. Cluster analysis of the major CPI-PPI network and the modules related to cell-cell signaling. Cluster 2 (A) is composed of 73 nodes and 354 edges, with $Ci = 4{,}849$. Cluster 2 contains the two hydrophilic substances cadmium and methylamine. Related GO term: regulation of cell communication. Cluster 18 (B) is composed of 13 nodes and 30 edges, with $Ci = 2, 304$. Cluster 18 contains one hydrophilic compound, trimethylamine, and one lipophilic compound, ethylbenzene. Related GO term: cell-cell signaling.
doi:10.1371/journal.pone.0061743.g004

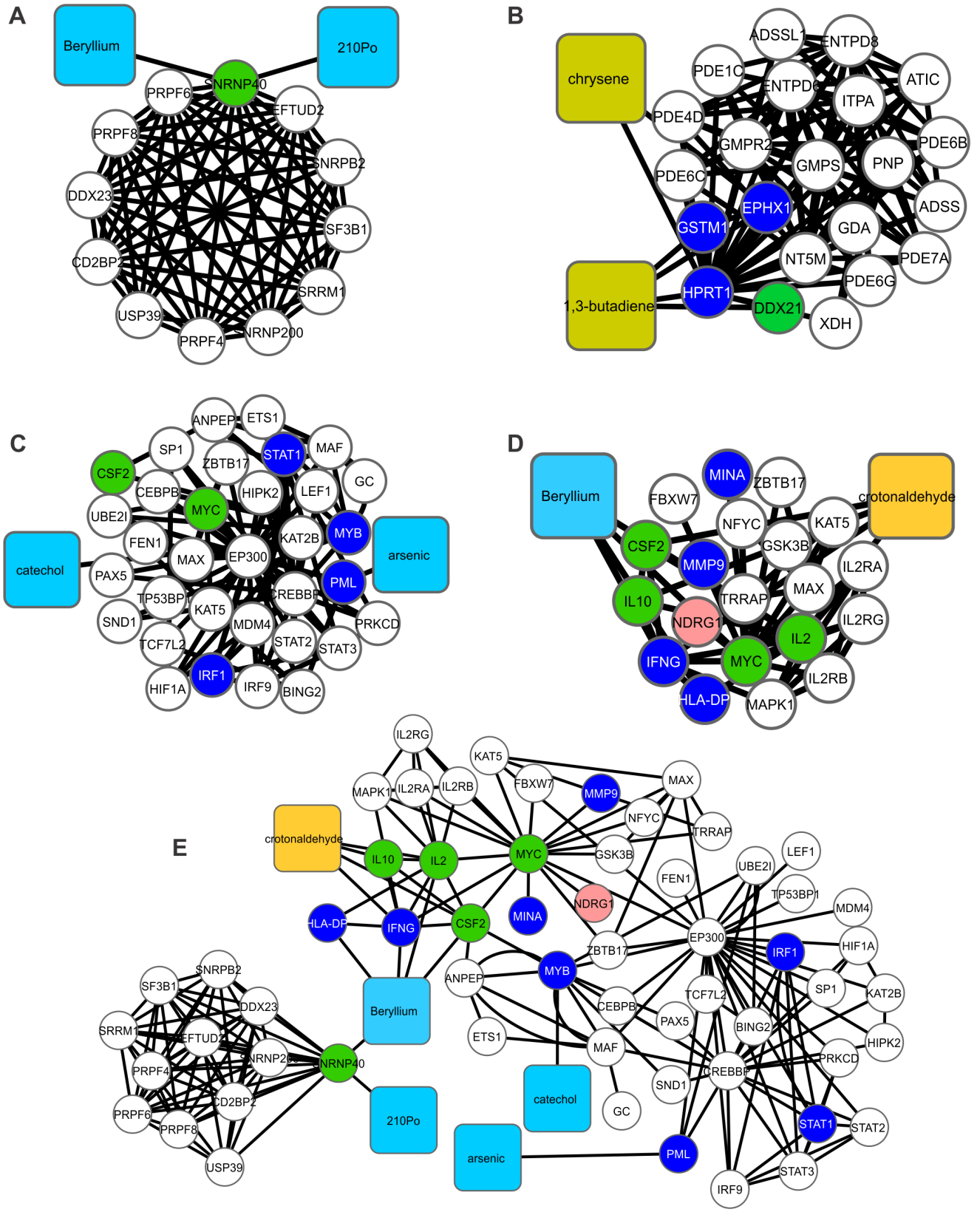

Figure 5. A merge of clusters 3, 6, 17 and 21. In (A), cluster 3 is composed of 14 nodes and 65 edges, with *Ci* = 4,643. The associated hydrophilic components are urethane, beryllium and polonium-210. Related GO terms: RNA-splicing and nucleobase, nucleoside, nucleotide and nucleic acid metabolic processes. Cluster 6 (B) is composed of 24 nodes and 102 edges, with *Ci* = 4,250. The associated lipophilic constituents are 1,3-butadiene and chrysene. Related GO term: nucleobase, nucleoside and nucleotide metabolic processes; cluster 17 (C) is composed of 35 nodes and 83 edges, with *Ci* = 2,371. The hydrophilic constituents present include catechol and arsenic. Related GO term: regulation of nucleobase, nucleoside nucleotide

and nucleic acid metabolic processes. Cluster 21 (D) is composed of 22 nodes and 49 edges, with $Ci = 2{,}227$. The associated hydrophilic constituent is beryllium, and the lipophilic constituent is crotonaldehyde. Related GO term: regulation of DNA metabolic processes. The union of clusters 3, 17 and 21 (E). Cluster 6 did not show any proteins overlapping with any other cluster.
doi:10.1371/journal.pone.0061743.g005

Results and Discussion

Data Prospecting and Topological Design of a Major CPI-PPI Network of Different Tobacco Constituents

Systems chemo-biology tools allow interactome networks of high-throughput data to be designed for CPI and PPI networks. In this sense, systems chemo-biology and systems pharmacology tools have been employed in different research areas, like prospection of new anticancer drugs [29], in order to evaluate the interaction of different small molecules with proteins and the main biological pathways potentially affected by these compounds.

Initially, our analysis was based on a list containing 95 TCs, extracted from [2]. From this initial list, we excluded compounds such as carbon monoxide, nitric oxide and phenol, which have different pleiotropic effects within a cell and could lead to the overrepresentation of many biological pathways not directly linked to development. In addition, we excluded all compounds without any protein target described, resulting in a final list containing 51 TCs commonly found in the mainstream and sidestream tobacco smoke (Table S2 in Supporting Information S1).

We have examined the relationship between 51 TCs and embryonic development pathways using systems chemo-biology tools. It should be noted that many of the thousands of substances in tobacco smoke are considered to represent public hazards, and some have carcinogenic potential [2]. Despite the growing interest in the elucidation of molecular pathways that can be affected by these compounds, many TCs do not have a known molecular target in the cell. However, our selected list of 51 TCs represents those substances with well described concentration in tobacco smoke, making them particularly attractive for experimental hypothesis testing. Moreover, these 51 TCs have some type of interaction with proteins already described, allowing systems chemo-biology studies. From this initial list of 51 TCs, we generated 51 small CPI-PPI networks (data not shown). Both STRING and STITCH add the nodes with the highest probability to be connected to a given node. Therefore, to create different CPI-PPI networks, we identified 20 to 50 additional proteins linked to each compound using only STITCH and STRING data and merged all of the networks using the Advanced Network Merge tool, which generates a single large network (referred to as the "main network", **Fig. 1A**). After creating the small networks, we found that RA receptors were present in the nicotine network. We decided to expand the nicotine network by adding a small network including RA and proteins related to RA signaling and embryonic development (**Fig. 1B**). The nicotine module was extracted from the first network to be studied independently because it showed a distinct module within the main network.

The resultant network after the nicotine module was extracted was referred to as the "major CPI-PPI network" and was composed of 898 nodes and 3,452 edges (**Fig. 1C**). It should be noted that, after merging each of the small CPI-PPI networks, two substances, 3-aminobiphenyl and dicyclohexyl, did not display any proteins in common with other compounds and were excluded from the analysis. Remarkably, the major CPI-PPI network did not show a wide overlap among the nodes, which indicates that TCs may have a broad influence and most likely affect different bioprocesses.

We next aimed to strengthen our understanding of our networks. We examined two types of data: (i) transcriptome data for each node directly associated with TCs to clarify whether the mRNA and, by inference, the proteins were present in the fetus (pink color), embryo (green color) or both (blue color) (Fig. 1, Table S1 in Supporting Information S1); and (ii) solubility predictions for the TCs and how this factor may influence the developing organism by characterizing each TC as hydrophilic or lipophilic (hydrophobic) (Fig. 1, Table S2 in Supporting Information S1). Nodes that did not show expression were left with uncolored (white) (Fig. 1, Table S1 in Supporting Information S1).

Interestingly, the majority of the nodes (145 of 234 total nodes; Fig. 1) have some role in human embryonic development, and thus, may affect the development of the organism. To predict TC solubility, we used the program ALOGPS 2.1. Among 48 TCs in our major CPI-PPI network (**Fig. 1**), we identified 21 lipophilic compounds and 27 hydrophilic components. Of the 27 hydrophilic components, 10 are inorganic, and 17 are organic (Table S2 in Supporting Information S1).

In addition, we used the program CentiScaPe 1.2 to examine the major CPI-PPI network for the most relevant proteins/compounds (**Figs. 1** and **2**). In a scale-free biological network, the most important nodes are the so called hub-bottlenecks (HBs) [30] because they combine the bottleneck function (nodes that controlling the information flow in a given network and displaying a betweenness score above the network average) and property hubs (nodes with a number of connections above the average node degree value of the network). Thus, HBs are critical nodes in a biological network [30]. In our analysis, we observed 143 HB nodes, of which 30 are TCs, and 53 were marked as present in both the fetus and embryo, 17 only in the embryo, 7 only in the fetus and 36 in neither the fetus nor embryo (white nodes) (**Fig. 2**). White nodes present in all of the networks are either not connected directly with the selected compound or do not show expression in any of the selected tissues. Because we only colored the direct nodes associated with a TC, it is clear that the TCs have a broad impact during development, acting in critical nodes that are necessary for development.

Furthermore, we sought to evaluate which TCs have the broadest effects on the major CPI-PPI network. Therefore, a closeness analysis was performed. Considering that the nodes showing the highest closeness are most relevant to the greatest number of nodes in a network [25], it can be assumed that the TCs exhibiting the highest closeness are those with the greatest systemic effects and impact the greatest number of proteins. A graph of closeness and betweenness was generated, showing that 33 TCs (from a total of 48) present closeness value above the average closeness of the network (Fig. S1 in Supporting Information S2). This finding is consistent with our interpretation that TCs have a systemic effect, impacting different proteins and physiological processes.

To understand how TCs interact with their targets, we analyzed the major CPI-PPI network for modules. From these analyses, we obtained the major TCs that affect different modules. After extraction of the nicotine subnetwork, MCODE found 22 significant modules (**Figs. 3–7**). Once the modules were obtained (**Figs. 3–7**), a gene ontology (GO) analysis was performed. Biological processes that are important for the development of organisms were listed (Table S3 in Supporting Information S1). Likewise, we performed additional GO analyses for the selected HBs (**Table 1**) and in each cluster (Tables S4–S25 in Supporting

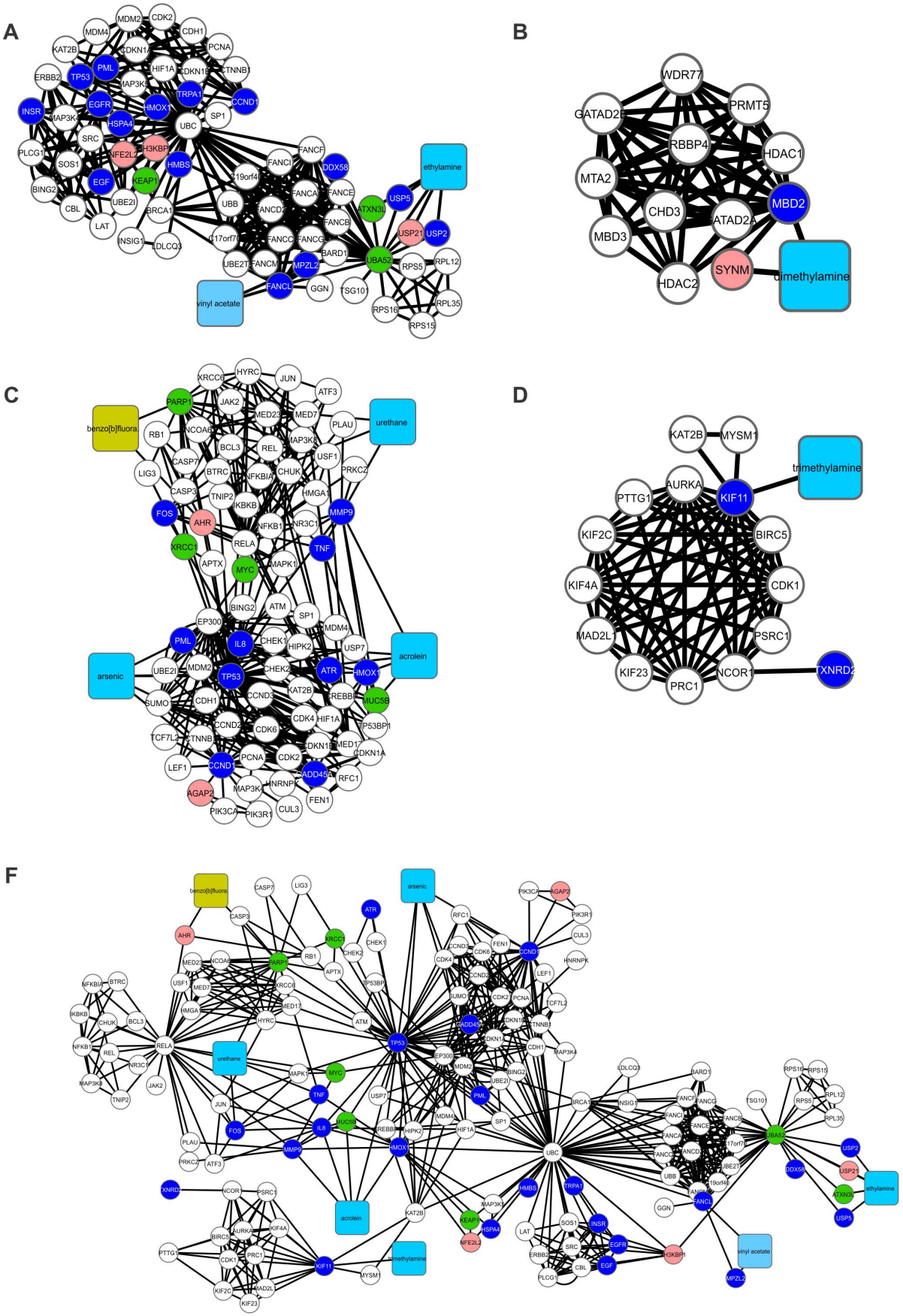
A
MDM2
CDK2
MDM4
CDH1
KAT2B
CDKN1A
PCNA
ERBB2
TP53
PML
HIF1A
CDKN1B
CTNNB1
MAP3K1
TRPA1
CCND1
INSR
EGFR
HMOX1
MAP3K4
HSPA4
SP1
UBC
SRC
PLCG1
SOS1
NFE2L2
H3KBP1
HMBS
FANCF
FANCI
DDX58
EGF
KEAP1
C19orf40
FANCA
FANCE
ethylamine
BING2
UBE2I
CBL
BRCA1
UBB
FANCD2
USP5
ATXN3L
LAT
FANCB
INSIG1
LDLCQ3
C17orf70
FANCC
FANCG
USP21
USP2
BARD1
UBE2T
FANCM
MPZL2
UBA52
RPS5
RPL12
vinyl acetate
FANCL
GGN
TSG101
RPS16
RPL35
RPS15
B
WDR77
GATAD2B
PRMT5
RBBP4
HDAC1
MTA2
CHD3
MBD2
ATAD2A
MBD3
SYNM
HDAC2
dimethylamine
C
HYRC
KRCC6
JUN
ATF3
JAK2
MED23
MED7
PARP1
PLAU
urethane
benzo[b]fluora
NCOA6
RB1
MAP3K8
BCL3
REL
USF1
CASP7
LIG3
BTRC
PRKCZ
CHUK
NFKBIA
CASP3
TNIP2
HMGA1
KBKB
FOS
AHR
NFKB1
NR3C1
MMP9
XRCC1
RELA
TNF
APTX
MYC
MAPK1
BING2
ATM
SP1
EP300
MDM4
PML
IL8
CHEK1
HIPK2
USP7
acrolein
arsenic
UBE2I
MDM2
CHEK2
ATR
HMOX1
TP53
CCND3
KAT2B
TREBB2
MUC5B
SUMO
CDH1
CDK4
HIF1A
CCND2
CDK6
TP53BP1
TCF7L2
CTNNB1
CDKN1B
MED1
CDKN1A
PCNA
CDK2
LEF1
CCND1
ADD45A
RFC1
MAP3K4
HNRNPK
AGAP2
FEN1
PIK3CA
PIK3R1
CUL3
D
KAT2B
MYSM1
trimethylamine
AURKA
KIF11
PTTG1
KIF2C
BIRC5
KIF4A
CDK1
MAD2L1
PSRC1
KIF23
PRC1
NCOR1
TXNRD2
F

Figure 6. Subnetworks derived from the merge of clusters 5, 8 and 9. In (A), Cluster 5 is composed of 69 nodes and 315 edges, with $Ci = 4{,}565$. The associated hydrophilic components are vinyl acetate and ethylamine. Related GO terms: response to DNA-damage stimulus and cell cycle. cluster 7 (B) is composed of 13 nodes and 54 edges with $Ci = 4{,}154$. The associated hydrophilic compound is dimethylamine. Related GO term: chromatin organization. Cluster 8 (C) is composed of 85 nodes and 338 edges with $Ci = 3{,}976$. The associated lipophilic constituents are acrolein and benzo[b]fluoranthene, whereas the hydrophilic constituents are urethane and arsenic. Related GO terms: DNA-damage stimulus and regulation of cell cycle. Cluster 9 (D) is composed of 16 nodes and 58 edges, with $Ci = 3{,}625$. The hydrophilic constituent present is trimethylamine. Related GO term: cell cycle processes. The union of clusters 5, 8 and 9 (F). Clusters 7 did not show any proteins overlapping with any other cluster.
doi:10.1371/journal.pone.0061743.g006

Information S1). Clusters that were not associated with significant GO terms due to a lack of data or were highly speculative in our analysis were excluded (Tables S13, S15 to S18, S22 and S25 in Supporting Information S1, Fig. S2 in Supporting Information S2).

Systemic Effects of Tobacco Smoking in Human Embryogenesis: Redox and Prostaglandin Metabolic Processes

The modularity data gathered from the major PPI-CPI network (**Fig. 1C**) were subjected to GO analysis. The GO analysis of clusters 1, 4, 11, 16, and 20 (**Fig. 3A–E**) revealed five main process annotations: (i) oxidation-reduction (redox), (ii) prostaglandin metabolism, (iii) steroid biosynthesis, (iv) lipid modification, and (v) unsaturated fatty acid metabolism (Tables S4, S7, S14, S19 and S23 in Supporting Information S1). Given the overlap among the different processes, these subnetworks were merged into a single network (**Fig. 3F**). It was observed that lipophilic molecules (e.g., chrysene, toluene, benz[a]anthracene, benzo[b]fluoranthene, 7H-dibenzo(c,g)carbazole, 2-naphthylamine, 4-aminobiphenyl and 5-methylchrysene; Table S2 in Supporting Information S1) were observed to be most connected to the proteins annotated as being involved in redox processes (**Fig. 3A**). Tobacco consumption has been associated with altered redox mechanisms and the generation of oxidative stress, leading to an inflammatory response [31], [32], [33], [34]. In this sense, within the merged network (**Fig. 3F**), two prostaglandin synthases (PTGS1 and PTGS2), and two 5-lipooxygenases (ALOX5 and ALOX15B), which play a role in the synthesis of leukotrienes [35], were identified. PTGSs are not only related to inflammatory responses when they are present at high levels in tissues but are also associated with normal pregnancies due to promoting adequate circulatory adaptation and regular maternal-fetal blood flow [32], [36]. In addition to the results of our GO analyses, it is known that maternal smoke diminishes prostaglandin levels, which causes low birth weight [36]. In addition, arsenic (**Fig. 3B**), which is present in this module, is related to increased oxidative stress via redox mechanisms [37]. Considering the data amassed in this module, it is possible to speculate that pro-oxidative stimulation by TCs, such as those included in **Fig. 3,** can generate a pro-inflammatory cascade, followed by downregulation of PTGSs and increased availability leukotriene, which promotes a continuous pro-inflammatory process. To corroborate this information, we used the transcriptomic data available for placenta and cord blood of passive smoking and non-smoking women [11]. In fact, the transcriptomic data analysis of placenta and cord blood of passive smoking women using landscape evaluation of the clusters 1, 4, and 20 (Fig. S1 in Supporting Information S3) indicated that the PTGS and ALOX genes are underexpressed when compared to non-smoking women. Interestingly, almost all glutathione S-transferase genes (e.g., GSTM1, GSTA1), which catalyze the conjugation of reduced glutathione with toxic xenobiotic substrates and confer antioxidative stress protection [38], are also downregulated in the placenta and cord blood of passive smoking women (Fig. S1 in Supporting Information S3), supporting the idea that TCs induce a pro-oxidative condition in embryo.

Systemic Effects of Tobacco Smoking on Human Embryogenesis: Regulation of Cell Communication and Cell-cell Signaling

Cellular communication is of great importance for embryonic development, being essential to coordinate the different biochemical signals required to control cellular differentiation and migration. Interestingly, GO analysis of clusters 2 and 18 (**Fig. 4**) revealed two related processes: (i) regulation of cell communication and (ii) cell-cell signaling (Tables S5 and S21 in Supporting Information S1). Considering the different proteins found in cluster 2 (**Fig. 4A**), two nodes appear to be important TC targets: (i) signal transducer and activator of transcription 3 (STAT3), which is related to cell-cell signaling in stem cell cultures [39]; and (ii) colony stimulating factor receptor-β (CSF2RB), a CSF2 receptor molecule that is important for post-blastocyst embryonic development, embryo differentiation, and implantation [40]. Epidermal growth factor (EGF) and its receptor EGFR were also present in this subnetwork (**Fig. 4A**). EGFR is a plasma membrane glycoprotein that is necessary for implantation and epithelial differentiation as well as for cell signal transmission during embryogenesis [41], [42]. It should be noted that both EGF and EGFR were linked to cadmium and methylamine (**Fig. 4A**) in our systems chemo-biology data. Other growth factors, such as nerve growth factor (NGF) and transforming growth factor α (TGFA), are also present in cluster 2. It is possible that the selected constituents, cadmium and methylamine (**Fig. 4**), can play a negative role in cell-cell signaling via inhibition of growth factors and its receptors. Considering the transcriptomic data available for the placenta and cord blood of passive smoking women [11], we observed that EGFR gene and other cell-cell signaling-associated genes are downregulated when compared to non-smoking women (Figs. S2 and S3 in Supporting Information S3).

It should be noted that in the major CPI-PPI network, 1,3-butadiene is linked to HOXD13 (**Fig. 1C**), whose mutations are associated with abnormal limb length [43]. Considering that tobacco abuse can lead to limb aberrations in newborns [3], the HOXD cluster should be an interesting target with respect to understanding the effects of cigarette compounds during development. Moreover, cluster 2 (**Fig. 4A**) contains ERBB2, ERBB3 and ERBB4, which are all members of the tyrosine kinase family and show a similar structure to EGFR, which appears to be crucial for skeletal development [44], and are also downregulated in the placenta and cord blood of passive smoking women (Fig. S2 in Supporting Information S3).

Systemic Effects of Tobacco Smoking in Human Embryogenesis: Metabolism of DNA, DNA Damage Stimulus, the Cell Cycle and Chromatin Organization

In the GO analysis of clusters 3, 6, 17 and 21 (**Fig. 5**), we identified two related processes: (i) RNA-splicing and (ii) metabolism of nucleotides and DNA (Tables S6, S9, S20 and S24 in

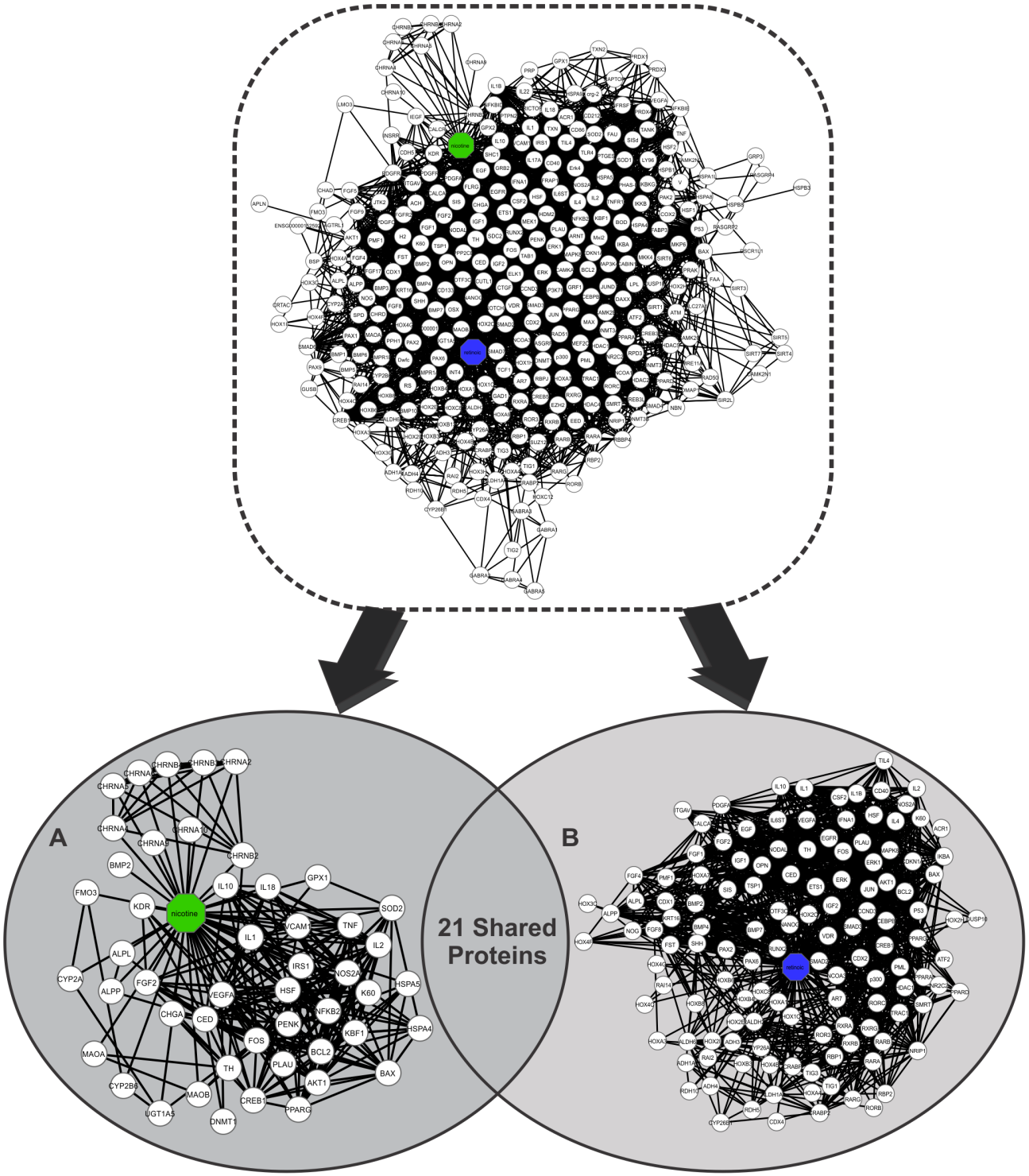

Figure 7. A binary network of the interactions between chemical compounds and proteins generated by the program Cytoscape 2.6.3, which contained 330 proteins and 4078 connections. Nicotine appears in the network as the green node, and RA appears as the blue node. White nodes are connected to both compounds are proteins. A) A subnetwork generated by the program Cytoscape containing 49 nodes and 281 edges and showing the proteins with direct connections with nicotine. B) A subnetwork generated by the program Cytoscape containing 130 nodes and 1,471 edges and showing the proteins that make direct connections with RA.
doi:10.1371/journal.pone.0061743.g007

Supporting Information S1). TCs were found associated with the metabolism of nucleotides in four different clusters, but each cluster contained different interacting compounds, including both hydrophilic (catechol) (**Fig. 5C to 5E**) and lipophilic (chrysene and 1,3-butadiene, crotonaldehyde) (**Fig. 5B**) as well as organic (chrysene and 1,3-butadiene) (**Fig. 5B**) and inorganic (beryllium,

Table 1. Major bioprocesses associated with the hub-bottleneck subnetwork.

GO-ID	GO	*p*-value	Corrected *p*-value	*k**	n#	Proteins
55114	Oxidation-reduction	4.4×10^{-16}	9.0×10^{-14}	26	645	CYP3A5;CYP1B1;PTGS2;CYP2C19;CYP2B6;PGD;PTGS1; ALDH3A2;AKR1C3;GSR;GPX1;FMO1;GPX4;HMOX1; GPX3;CAT;NQO1;HADH;CYP1A1;CYP2C8;MAOB;IDO1; CYP2E1;CYP1A2;DECR1;SOD2;LDLCQ3;ALDH7A1;G6PD; AKR1B1;TXN;ALDH2;MPO
48545	Regulation of steroid hormone stimuli	1.3×10^{-13}	1.8×10^{-11}	18	225	TNF;PTGS2;MAP2K1;RELA;PTGS1;MAOB;BRCA1; MAPK1;FOS;CCND1;CDKN1A;EP300;HMOX1;GPX4; GPX3;ALDH2;INSR;NGF
42127	Regulation of cell proliferation	1.3×10^{-12}	1.5×10^{-10}	30	848	CSF2;TNF;GNAI2;PTGS2;PTGS1;TAC1;GPX1;INS;HMOX1;IFNG;SHC1; EGF;INSR;MYC;EGFR;KAT2B;IL8;MAP2K1; RELA;TP53;IDO1;STAT1; MBD2;BRCA1;SOD2;MAPK1; CDKN1A;CCND1;IL2;NGF
42981	Regulation of apoptosis	9.3×10^{-12}	9.4×10^{-10}	29	282	CSF2;TNF;PTGS2;MMP9;GPX1;APP;INS;ALB;HMOX1; SOS1;IFNG;CAT;NQO1; EGFR;RELA;GRIN2A;TP53;IDO1; STAT1;BRCA1;SOD2;MAPK1;CDKN1A;F2;MPO;PDCD6; GSTP1;IL2;NGF
10646	Regulation of cell communication	6.0×10^{-10}	2.6×10^{-8}	31	1154	CSF2;TNF;GNAI2;PTGS2;CD8A;GRB2;TAC1;GPX1;APP; INS;SOS1;HMOX1;IFNG;CHRNA4;SHC1;CAT;EGF;INSR; AGAP2;EGFR;MAP2K1;RELA;MAOB;GRIN2A;TP53; MBD2;PTPN11;LAP3;CCND1;IL2;NGF

*Number of nodes for a given GO in the network;
#Total number of proteins for a given GO annotation.
doi:10.1371/journal.pone.0061743.t001

polonium-210 and arsenic) substances (**Fig. 5A, 5C, 5D, and 5E**). Remarkably, in cluster 6, these substances are linked to HPRT1 (**Fig. 5B**), a hypoxanthine phosphoribosyltransferase that is responsible for the metabolism of purines [45].

Moreover, 1,3-butadiene, has been found to be linked to increased genotoxic stress due to DNA damage through the formation of DNA-DNA cross-links at adenine and guanine nucleobases by its metabolites, 1,2,3,4-diepoxybutane and 3,4-epoxy-1,2-butanediol [46], [47]. The compound 1,3-butadiene has also been associated with epigenotoxic effects caused by the loss of global DNA methylation and trimethylation of histone H3 lysines 9 and 27 and H4 lysine 20, all of which are known for their roles in regulating gene expression patterns [47].

Next, in the GO analysis of clusters 5, 8 and 9 (**Fig. 6**), we identified two related processes: (i) DNA damage stimulation and (ii) the cell cycle (Tables S8, S11 and S12 in Supporting Information S1). In this cluster, arsenic binds directly to PLM (**Fig. 6B and 6D**), which is a protein with functions involved in chromatin organization, cell differentiation, DNA repair, protein sequestration and post-translational modifications [http://www.genecards.org]. PML is linked to significant proteins that regulate cell cycle such as p53, p300 and BING2 (**Fig. 6B and 6D**). Another TC, urethane, is directly connected to FOS (**Fig. 6B and 6D**), a central protein involved in proliferation, and TNF, a pro-inflammatory cytokine. Urethane is reported to alter placental morphology and down-regulates cell cycle genes as well as cytokines and other growth factors [48]. Interestingly, TCs downregulate the expression of genes associated with the metabolism of nucleotides and DNA, and cell cycle, as observed by transcriptomic analysis (Figs. S4 and S5 in Supporting Information S3).

In the GO analysis of cluster 7 (**Fig. 6B**), we only identified chromatin organization (Table S10 in Supporting Information S1) as a major biological process. Cluster 7 included dimethylamine, which is connected to MBD2 (**Fig. 6B**), a protein associated with regions of methylated DNA in CpG islands that can recruit histone deacetylases (HDACs) and DNA methyltransferases [http://www.genecards.org]. DNA methylation is also correlated with gene silencing through polycomb repression complexes (PRC) [49]. PRC is involved in the silencing of many HOX genes [49], which are critical for normal fetus development. Additionally, MBD2 is correlated with the inactivation of sexual chromosomes and is a candidate for recruiting DNA-methyltransferases (DNMTs) to the silenced promoters of long-term repressed genes [50]. Taking into account the effects of TCs in the expression of genes associated to chromatin remodeling, like HDACs, it can be observed that placenta and cord blood of passive smoking women showed a downregulation of those genes (Fig. S6 in Supporting Information S3), supporting the idea that TCs can affect chromatin remodeling during embryogenesis.

Effect of Nicotine on Retinoic Acid Signaling, Cell Proliferation and Differentiation

A second analysis using systems chemo-biology tools was developed to elucidate the relationships between nicotine, RA signaling and cell differentiation in the fetus during embryonic development in female smokers. The extracted subnetwork was examined separately due the distinct module involving nicotine and its interacting proteins. RA was added to the network because we observed that many proteins connected to nicotine are related to embryonic development and RA signaling.

Thus, the amassed data allowed the design of a major CPI network associated with nicotine and RA signaling (**Fig. 7**), which revealed several proteins that related to embryonic development, stress responses, and cell proliferation. Several of the proteins in the CPI network are directly connected to nicotine, including (i) VEGFA, a factor that induces blood vessel formation (angiogenesis) [51]; (ii) DNMT1, a DNA methyltransferase responsible for the methylation of 5′CpG islands in DNA (**Fig. 7**) [52]; (iii), FOS and JNK1 (MAPK8), which are both inducers of cell proliferation [53], [54]; and (iv) SOD2, which is responsible for mitochondrial superoxide dismutation. In addition, many proteins involved in cellular responses to stress, DNA damage and inflammation are interconnected with nicotine in the CPI network (**Fig. 7**).

We observed a connection between nicotine and JNK1 through their association with RARα in the CPI network (**Fig. 7**). JNK1 is

Table 2. The relationships between common proteins, nicotine and RA and their specific biochemical functions. These data were obtained from the GeneCards (http://www.genecards.org) and iHop (http://www.ihop-net.org/UniPub/iHOP/) databases.

Protein	Biological function	Role
VEGFA	Growth factor	Crucial role in angiogenesis, vasculogenesis and endothelial growth
TGFB1	Cytokine	Acts in differentiation, proliferation, adhesion and migration; also a potent stimulator of bone growth
ALPP	Alkaline phosphatase	Expressed in the placenta
HSF	Transcription factor	Activated under conditions of heat or other cellular stress
FGF2	Growth factor	Involved in tumor growth, development of the nervous system, cell differentiation and angiogenesis
TH	Hydroxylase	Hydroxylase that functions in the physiology of adrenergic neurons
FOS	Nuclear phosphoprotein	Nuclear phosphoprotein that participates in cell differentiation, proliferation and apoptosis
IL10	Cytokine	Involved in the immune response against pathogens and in the inflammatory response; also related to the intestinal immune system
DNMT1	Methyltransferase	DNA methylation and the establishment of methylation patterns
IL1	Cytokine	Involved in the immune response to pathogens and the inflammatory response
AKT1	Kinase	Involved in tumor formation, angiogenesis and insulin regulation
BAX	Transcription factor	Pro-apoptotic protein
ALPL	Alkaline phosphatase	Mineralization of bone matrix
BCL1 (IL5)	Cytokine	Involved in the immune response against pathogens and the inflammatory response
NOS2A	Nitric oxide synthase	Produces nitric oxide (NO)
PPARG	Proliferator peroxisome receptor	Regulator of adipocyte differentiation and glucose homeostasis
K60 (IL8)	Chemokine	Involved in the inflammatory response; angiogenesis inducer
CREB1	Transcription factor	Controls circadian rhythm, tumor suppressors and the expression of various genes involved in cell survival
PLAU	Protease	Involved in extracellular matrix degradation and possibly tumorigenesis
IL2	Cytokine	Essential in the proliferation of T-cells of the immune system. Stimulates the production of B-cells, monocytes and natural killer cells
KDR (VEGFR)	Growth factor	Plays a crucial role in vasculogenesis and angiogenesis
RARB	Retinoic Acid Receptor	Involved in cell differentiation, cell growth arrest, and signaling and transcription of target genes

doi:10.1371/journal.pone.0061743.t002

expressed when the cell undergoes cellular stresses, such as inflammation, oxidative stress, and heat [55]. In a murine model, nicotine was found to be related to the expression of JNK1 in respiratory system tissues through nicotinic receptors and receptor kinins B1 and B2, whose stimulation by bradykinin leads to increased levels of intracellular Ca^{2+} [53]. Cellular stress can activate JNK1, which phosphorylates RARα and causes its proteosomal degradation [55].

Supporting the idea that nicotine can induce the activation of pro-inflammatory cascades and different cellular stress pathways, the placenta and cord blood of passive smoking women showed an upregulation of interleukin receptors (e.g., IL2RA; IL2RB), VEGFA, FOS, JAK1, among others (Fig. S7 in Supporting Information S3). Moreover, genes associated with antioxidative stress, like SOD2, are underexpressed when compared to non-smoking women (Fig. S7 in Supporting Information S3).

The systems chemo-biology analysis performed in this study also showed that nicotine is directly connected to the protein CYP26A1 (**Fig. 7**), whose coding gene is downregulated in placenta and cord blood of passive smoking women (Fig. S7 in Supporting Information S3). This protein is responsible for regulating RA levels [56] and is expressed in a spatial-temporal manner during the development of mice, mainly in the anterior segment of the embryo and in the neural crest-derived mesenchyme [56]. However, inhibition of this protein generates an accumulation of RA and leads to deformities in the embryo, such as abnormalities in the cerebellum, urogenital tract, and spinal cord [56]. Moreover, nicotine exhibited 22 proteins in common with RA (**Table 2**). These proteins are mostly related to the immune system, stress, and cell proliferation (**Table 2**), indicating that nicotine affects RA signaling through cellular stress caused by constant tobacco use. Interestingly, we observed that nicotine was directly linked with VEGFA in our analysis (**Fig. 7**). Exposure to nicotine could results in an increase in pro-inflammatory signaling, leading to abnormal expression of VEGFA and other placental growth factors, reducing uroplacental blood flow and culminating in fetal growth restriction [57] (**Fig. 8**), an idea that is supported by transcriptomic data (Fig. S7 in Supporting Information S3).

Role of Nicotine in the Differentiation of Bone Tissue

An indirect association of nicotine with RA receptors was observed in the network via the influence of nicotine on the transcription factor JUN (**Fig. 7**). The JUN protein can be activated by the action of JNK1 during osteoblast differentiation [58]. In a smoking woman the blood concentration of nicotine are maintained at a stable level depending on the degree of tobacco use [59]. During embryogenesis, constant levels of nicotine can affect bone development, and morphological data have demon-

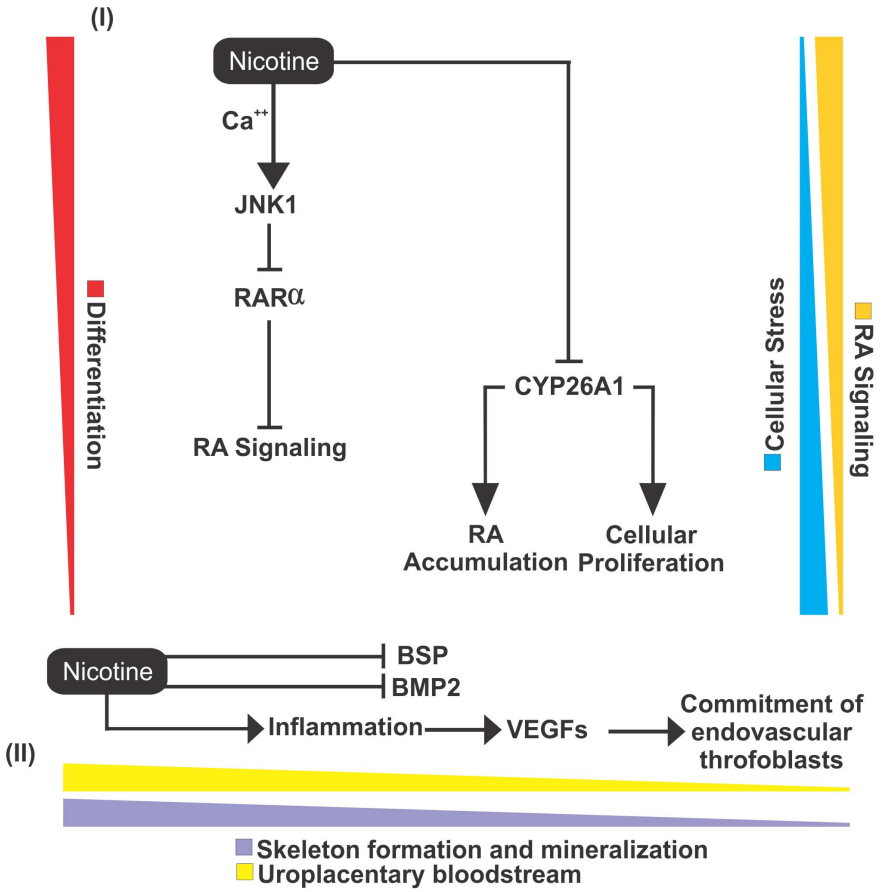

Figure 8. A molecular model illustrating how nicotine could potentially affects differentiation. In the first part of the model (I), it can be observed that by generating cellular stresses, nicotine promotes the recruitment of JNK1 through the influx of intracellular Ca^{2+}. JNK1, by itself, promotes the inhibition of RARα. Finally, nicotine promotes the inhibition of CYP26A1, which generates an accumulation of RA in the cell and an increase in cell proliferation. In the second part of the model (II), the inhibition of BMP2 and BSP is promoted by nicotine, which results in the negative regulation of bone mineralization and skeletal development. In addition, nicotine promotes a pro-inflammatory reaction that recruits VEGF and placental growth factors, which leads to an impairment of the endovascular trophoblast, resulting in a fetal growth restriction.
doi:10.1371/journal.pone.0061743.g008

strated a decrease in bone and cartilage growth [60]. An additional impact of nicotine on bone tissue differentiation involves the relationship with the BMP2 and BSP proteins. The BMP protein family includes the most potent osteogenic growth factors described to date [51] and is connected to nicotine (**Fig. 7**). A study in rabbits showed that treatment with nicotine affects BMP2 RNA levels and the activity of osteoblasts [51]. Similarly, the BSP protein is a glycoprotein that acts on bone mineralization, which has also been described as being inhibited by nicotine in rat osteoblast cells [61]. Corroborating these findings, the transcriptomic data of placenta and cord blood of passive smoking women support the fact that nicotine and other TCs inhibit the expression of BMP2 (Fig. S8 in Supporting Information S3).

Modularity and Centrality Analyses Linking Nicotine with Abnormal Embryonic Development

Once the CPI network was generated (**Fig. 7**), we aimed to understand which major protein clusters might be present. In this sense, the CPI network (**Fig. 7**) showed the presence of six modules with a coefficient of cohesion greater than or equal to 3.00 (Clusters 1–6, Fig. S3 in Supporting Information S2). It was observed that nicotine appeared in clusters 1–4 (Fig. S3A–D in Supporting Information S2), but not associated with RA (only in Fig. S3C in Supporting Information S2), which exhibits many connections other than nicotine in the network. Nicotine is connected to 49 proteins with 281 connections, and RA is connected to 130 proteins with 1,471 connections (**Fig. 7**). From the systems chemo-biology analysis, it was observed that nicotine more readily clustered in a network focused on proteins involved in development and cellular stress (**Fig. 7**). We also observed that certain clusters did not contain either nicotine or RA (Figs. S3E and F in Supporting Information S2). In cluster 5 (Fig. S3E in Supporting Information S2) there are a prevalence of proteins linked to (i) chromatin remodeling, such as EZH2, EED, SUZ12, DNMT1, DNMT3A, DNMT3B, HDAC2, HDAC4, HDAC5, and (ii) development and differentiation, including several HOX proteins [A1, 1C (A5), 4B (D4), 2I (B1), B13, B4 and 4F (A11)], PAX1, PAX6, NANOG, RAR, RXRβ, NOTCH1, CYP2B6, and CYP26A1.

To identify the major nodes within the CPI network (**Fig. 8**), we calculated betweenness, closeness and node degree centralities. From these analyses, two graphs were generated containing the proteins that showed the highest centrality values (Figs. S4 and S5 in Supporting Information S2). Interestingly, these nodes present a similar relevance order in both graphs. Thus, RA, nicotine, JNK1, p300, AKT1, p53 and ERK showed the highest betweenness, closeness and node degree values (Figs. S4 and S5 in Supporting

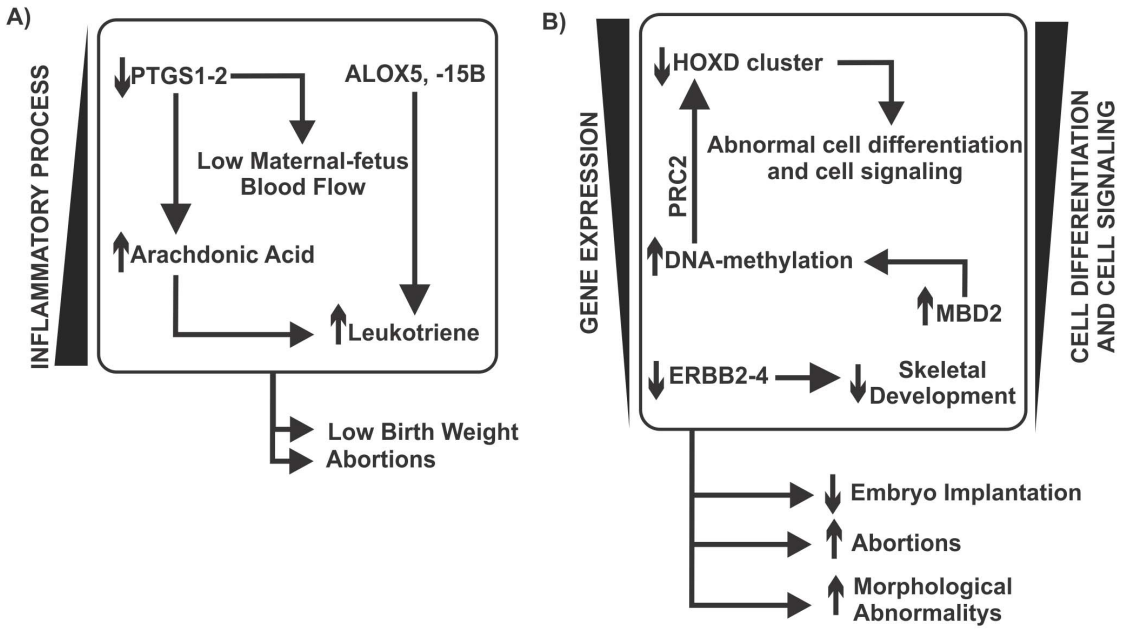

Figure 9. A model of the interactions from a systemic view showing how TCs affect development. In (A), we show that increasing TC levels generate a pro-inflammatory cascade by increasing the levels of PTGS1 and PTGS2. PTGSs are associated with inflammatory responses and are essential for normal pregnancy. Disturbances in PTGS expression could cause impairments in fetal development. TCs are connected to ALOX5 and ALOX15B, which are proteins involved in the synthesis of leukotriene, a molecule that plays pivotal roles in pro-inflammatory responses. The consequence of (A) is low birth weight in newborn infants, abortions and increased proliferation. Moreover, in (B), TCs are linked to BING2 and USP2, which are proteins related to increased activity of MDM2. This MDM2 mediated up-regulation can rapidly down-regulate p53 protein, leaving the cell more susceptible to DNA damage. TCs also down-regulate HPRT1, diminishing purine metabolism. This system exhibits a relationship with increased proliferation. The systems in (C) shows that TCs are associated with the generation of superoxides due to up-regulating NADH oxidase, which increases ROS levels and, consequently, oxidative stress. Increased oxidative stress is known to be related to birth defects. In addition, system (C) is associated with low birth weight and low neutrophil activity. Moreover, (D) shows the relationship between TCs and low hormone synthesis and signaling. Exposure to TCs could have a negative effect on androgen and estrogen solubility due to acting on the UGT cluster. TCs could also be associated with low levels of cholesterol synthesis due to increasing the levels of CYPs and diminishing the levels of FDFT1 and FDPS, which are two enzymes related to cholesterol synthesis. Low cholesterol availability would decrease general hormone synthesis. In addition, TCs affect the transport of cholesterol to the mitochondria by acting on the membrane protein StAR. Finally, in (E), the system shows the relationships between TCs and decreased global gene expression and cellular differentiation and signaling. The activities of the TCs would increase the levels of MBD2, a methylation enzyme. DNA methylation is related to gene silencing. We postulate that TCs could affect the PRC2 complex via its methylation and disturb gene expression, including that of HOX genes. TCs could also have a negative effect on gene expression by increasing YWHAH levels, which would decrease the levels of the master kinase PDPK1 and is linked to AKT activation and SMAD nuclear translocation. In addition, NOTCH signaling could be affected through the action of TCs on APP activation.
doi:10.1371/journal.pone.0061743.g009

Information S2). As these proteins play major roles in cellular physiology, it was expected that they would exhibit higher values for the three variables. The proteins with the highest values were taken into consideration in the design of a molecular model of the effect of nicotine on embryonic development (**Fig. 8**). In the centrality analysis, it was observed that p300 appeared as an important node, showing the highest values of betweenness, closeness, and the node degree (Figs. S4 and S5 in Supporting Information S2). This scenario demonstrates that there is a major influence of p300 on the network regarding the number of connections with other proteins (92 proteins), the implications of its importance for neighboring proteins (closeness) and its relationships to clusters and bioprocesses (betweenness). Therefore, the negative regulation of this protein induced by nicotine can also lead to fetal malformations and could be a potential study target for understanding the influence of nicotine in development. Noteworthy, the transcriptomic analysis of extraembryonic tissues extracted from pregnant passive smoking women showed a downregulation of p300-coding gene (Fig. S8 in Supporting Information S3).

An important issue that should be addressed in the future is the influence of the major nicotine metabolites on the activity of the enzymes and proteins observed in this work. It has been reported that 70–80% of nicotine is metabolized to cotinine by CYP2A6 to produce nicotine and a cytoplasmic aldehyde oxidase [62]. However, nicotine can generate an elevated number of different metabolites, whose mechanism of action is not clear [62]. Additionally, the mechanism of detoxification of nicotine and cotinine is based on the glucuronidation of both molecules, accounting for 40–60% of the nicotine found in urine [62]. Unfortunately, for the majority of compounds present in tobacco smoke observed in this work, the data about its metabolization or detoxification are virtually unknown. The use of metabolomic techniques associated with systems chemo-biology tools should improve our understanding of how nicotine and other TCs physiologically affect development.

Conclusions

In the present study, we showed, using systems chemo-biology tools, how the primary harmful constituents of tobacco interact with specific biological processes and affect them. Our cluster analysis results show that TCs act in many bioprocesses, including cell communication and signaling, hormone synthesis and

signaling, DNA metabolism, DNA repair, and inflammation, whose results were supported by landscape network analysis of transcriptomic data of extraembryonic tissues gathered from passive smoking women and non-smoking women. Although these processes have wide effects on cellular and embryonic physiology, they can be disturbed by the levels of the constituents of tobacco smoke. Because these effects are complex, we developed an interaction which comprises two main mechanisms associated with TCs: increased inflammatory processes **(Fig. 9A)**, and negative regulation of gene expression, cell differentiation and cell signaling (**Fig. 9B**). The systems model is related to low birth weight, an increased probability of abortion, morphological abnormalities (mainly in the skeletal system), low neutrophil activity and increased proliferation rates. Furthermore, our model can help improve knowledge and provide new insights regarding how the chemicals in tobacco cause the many morphological abnormalities observed in the newborn offspring of smoking pregnant women. The role of nicotine in embryonic development has also not been well studied. The analysis performed in this study demonstrates that nicotine has an aggressive effect on cell differentiation, affecting RA signaling in the embryo, inhibiting RA receptors due to intracellular calcium influx and stimulating cell proliferation proteins that antagonize RA activity. Osteoblast differentiation is also affected by nicotine via inhibiting proteins that stimulate bone tissue formation, which complements the TC model. Together, these data show that the birth defects observed in morphological studies could be caused by the negative action of nicotine on RA signaling. The networks also show that the pro-inflammatory pathway triggered by nicotine could be a factor leading to decreased body weight in the fetuses of smoking women. Finally, cluster analysis shows a systemic effect of nicotine, which could affect the network in a more aggressive and short-term way via cellular stress cascades.

Supporting Information

Supporting Information 1 **Table S1** Transcriptomic data of the proteins directly linked to the selected tobacco constituents (TCs). **Table S2** List of tobacco constituents (TCs) found in the major CPI-PPI network (Fig. 1). The solubility of each compound was accessed using the program ALOGPS 2.1. Those compounds with solubility less than 20 g/l were considered lipophilic. **Table S3** GO processes present in the main tobacco constituents (TCs)-associated CPI-PPI network (Fig. 1). **Table S4** GO processes present in the cluster 1 (Fig. 3A). **Table S5** GO processes present in the cluster 2 (Fig. 4A). **Table S6** GO processes present in the cluster 3 (Fig. 5A). **Table S7** GO processes present in the cluster 4 (Fig. 3B). **Table S8** GO processes present in the cluster 5 (Fig. 6A). **Table S9** GO processes present in the cluster 6 (Fig. 5B). **Table S10** GO processes present in the cluster 7 (Fig. 6B). **Table S11** GO processes present in the cluster 8 (Fig. 6C). **Table S12** GO processes present in the cluster 9 (Fig. 6D). **Table S13** GO processes present in the cluster 10 (S-Fig. 2A). **Table S14** GO processes present in the cluster 11 (Fig. 3C). **Table S15** GO processes present in the cluster 12 (S-Fig. 2B). **Table S16** GO processes present in the cluster 13 (S-Fig. 2C). **Table S17** GO processes present in the cluster 14 (S-Fig. 2D). **Table S18** GO processes present in the cluster 15 (S-Fig. 2E). **Table S19** GO processes present in the cluster 16 (Fig. 3D). **Table S20** GO processes present in the cluster 17 (Fig. 5C). **Table S21** GO processes present in the cluster 18 (Fig. 4B). **Table S22** GO processes present in the cluster 19 (S-Fig. 2F). **Table S23** GO processes present in the cluster 20 (Fig. 3E). **Table S24** GO processes present in the cluster 21 (Fig. 5D). **Table S25** GO processes present in the cluster 22 (S-Fig. 2G).
(XLSX)

Supporting Information 2 **Figure S1** Graph showing the relationship of closeness and betweenness of the TCs in the major CPI-PPI network. All nodes in the graph present a mean above average in both closeness and betweenness. The color represents the soluble property of the TCs (Light blue = hydrophilic and Yellow = lipophilic). Three nodes have distinct color/shape, since they shared a color with the adjacent node [Chromium = Large width node (black); Cadmium = Diamond shape/blue colored; and 7H-dibenzo[cg]carbazole = Orange node]. **Figure S2** Clusters excluded from the analysis due lack of literature data associated with TCs and their given GO, therefore, being highly speculative. In (A), Cluster 10 is composed by 12 nodes and 39 edges, with Ci = 3,250. The associated hydrophilic component is furfural. Related GO: Glucose Catabolic Process and Pentose-Phosphate Shunt. Cluster 12 (B) is composed by 16 nodes and 43 edges, with Ci = 2,750. The associated hydrophilic components are cadmium and acrynolitryle. Related GO: Antigen Processing and Presentation. Cluster 13 (C) is composed by 18 nodes and 48 edges, with Ci = 2,667. The associated hydrophilic component is urethane and the lipophilic is xylene. Related GO: G-Protein Coupled Receptor Protein Signaling Pathway. Cluster 14 (D) is composed by 42 nodes and 109 edges, with Ci = 2,595. The associated hydrophilic components are hydrazine, resorcinol, nickel and chromium. Related GO: Regulation of Insulin Signaling Pathway. Cluster 15 (E) is composed by 22 nodes and 55 edges, with Ci = 2,250. The associated hydrophilic components are chromium and acrynolitryle. Whereas the lipophic are xylene, chrysene, 5-methylcrysene, benz[a]anthracene and benzo[b]fluoracene. Related GO: Response to Chemical Stimuli. Cluster 19 (F) is composed by 12 nodes and 27 edges, with Ci = 2,250. The associated hydrophilic component is lead. Related GO: I-KappaB Kinase/NF-KappaB Cascade. Cluster 22 (G) is composed by 20 nodes and 43 edges, with Ci = 2,150. The associated hydrophilic components are cadmium, lead, pyrrole and arsenic. Related GO: Heme Biosynthetic Process. **Figure S3** Clusters 1 to 6, extracted from the nicotine CPI-PPI network by MCODE. The blue node is RA and the green node is nicotine. Cluster 1 (A) is composed by 159 nodes and 2373 edges, with Ci = 14, 925; Cluster 2 (B) is composed by 227 nodes and 2649 edges, with Ci = 11,670; Cluster 3 (C) is composed by 207 nodes and 1793 edges, with Ci = 8,662; Cluster 4 (D) is composed by 174 nodes and 1002 edges, with Ci = 5,759; Cluster 5 (E) is composed by 89 nodes and 300 edges, with Ci = 3,371; Cluster 6 (F) is composed by 12 nodes and 36 edges, with Ci = 3,000. Nicotine appears in four clusters (A to D), whereas RA only in C, showing that nicotine is more easily clustered. **Figure S4** Graph showing the relationship of node degree (ND) and betweenness (BT) using all proteins in the nicotine CPI-PPI network. The seven most significant proteins were selected (which are present near the value of 5.0×10^3). The dotted line shows the threshold of significance, and the values above the line are considered more relevant. **Figure S5** Graph showing the relationship of closeness (CL) and betweenness (BT) from all proteins in the nicotine CPPI-PPI network. The seven most significant proteins were selected (which are present near the value of 5.0×10^3). The dotted line shows the threshold of significance, and the values above the line are considered more relevant.
(DOCX)

Supporting Information 3 **Figure S1** Network representation of cluster 1,4, and 20 obtained from STRING metasearch engine

(A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoking women (a) versus placenta plus cord blood from non-smoking women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S2** Network representation of cluster 2 obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoking women (a) versus placenta plus cord blood from non-smoking women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S3** Network representation of cluster 18 obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoking women (a) versus placenta plus cord blood from non-smoking women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S4** Network representation of cluster 3, 11 and 21 obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoker women (a) versus placenta plus cord blood from non-smoker women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S5** Network representation of cluster 5, 8 and 9 obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoker women (a) versus placenta plus cord blood from non-smoker women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S6** Network representation of cluster 7 obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoker women (a) versus placenta plus cord blood from non-smoker women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S7** Nicotine-associated network obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoker women (a) versus placenta plus cord blood from non-smoker women (b)]. In this sense, the mathematical equation z = a/(a+b) was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when z >0.55 (yellow to red colors), lower than (b) when z <0.45 (cyan to blue colors) and equivalent to (b) when 0.45< z <0.55 (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as ''3D-Graph'', build on ''node'', resolution ''level-50'', contrast ''level-50'', smoothness ''level-50'' and zoom ''level-50''. **Figure S8** Retinoic acid-associated network obtained from STRING metasearch engine (A). This network was used for two-state landscape analysis of gene expression (B). Coordinates (X- and Y-axis) represent normalized values of the input network topology. Color gradient (Z-axis) represents the relative gene functional state mapped onto network according to the transcriptomic data input

of GSE30032 series file [placenta plus cord blood transcriptomic data from passive smoker women (a) versus placenta plus cord blood from non-smoker women (b)]. In this sense, the mathematical equation $z = a/(a+b)$ was used to calculated the relative gene functional state of condition (a) and condition (b). Thus, the gene expression in condition (a) is greater than condition (b) when $z > 0.55$ (yellow to red colors), lower than (b) when $z < 0.45$ (cyan to blue colors) and equivalent to (b) when $0.45 < z < 0.55$ (green color). The landscape was generated by ViaComplex 1.0 software with the following options: plot as "3D-Graph", build on "node", resolution "level-50", contrast "level-50", smoothness "level-50" and zoom "level-50".
(DOCX)

Author Contributions

Conceived and designed the experiments: BCF JFP DB. Performed the experiments: BCF JFP DB. Analyzed the data: BCF DB. Contributed reagents/materials/analysis tools: BCF JFP DB. Wrote the paper: BCF JFP DLN DB.

References

1. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, et al. (2002) Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene 21(48): 7435–7451.
2. Fowles J, Bates M (2000). The chemical constituents in cigarette and cigarette smoke: priorities for harm reduction. Available: http://www.moh.govt.nz/moh.nsf/pagescm/1003/$File/chemicalconstituentscigarettespriorities.pdf. Accessed 2013 March 26.
3. Hackshaw A, Rodeck C, Boniface S (2011) Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. Human Reprod Update 17(5): 589–604.
4. Florescu A, Ferrence R, Einarson T, Selby P, Soldin O, et al. (2009) Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology. Ther Drug Monit 31(1): 14–30.
5. Morris CV, DiNieri JA, Szutorisz H, Hurd YL (2011) Molecular mechanisms of maternal cannabis and cigarette use on human neurodevelopment. Eur J Neurosci 34(10): 1574–1583.
6. Sadeu JC, Foster WG (2011) Cigarette smoke condensate exposure delays follicular development and function in a stage-dependent manner. Fertil Steril 95(7): 2410–2417.
7. Jauniaux E, Burton GJ (2007) Morphological and biological effects of maternal exposure to tobacco smoke on the feto-placental unit. Early Hum Dev 83(11): 699–706.
8. Duester G (2008) Retinoic acid synthesis and signaling during early organogenesis. Cell 134(6): 921–931.
9. Daftary GS, Taylor HS (2006) Endocrine regulation of HOX genes. Endocr Rev 27(4): 331–355.
10. Cheng CG, Lin B, Dawson MI, Zhang XK (2002) Nicotine modulates the effects of retinoids on growth inhibition and RARβ expression in lung cancer cells. Int J Cancer 99 (2): 171–178.
11. Votavova H, Dostalova M, Krejcik Z, Fejglova K, Vasikova A, et al. (2012) Deregulation of gene expression induced by environmental tobacco smoke exposure in pregnancy. Nicotine Tob Res 14(9): 1073–1082.
12. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, et al. (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37: D412–D416.
13. Snel B, Lehmann G, Bork P, Huynen MA (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acid Res 28(18): 3442–3444.
14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504.
15. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D (1997) GeneCards: integrating information about genes, proteins and diseases. Trends Genet 13: 163.
16. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, et al. (2010) GeneCards Version 3: the human gene integrator Database. Database (Oxford) 2010: baq020.
17. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27–30.
18. Hoffmann R, Valencia A (2004) A gene network for navigating the literature. Nature Genetics 36: 664.
19. Tetko IV, Tanchuk VY (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. J Chem Inf Comput Sci 42(5): 1136–1145.
20. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, et al. (2009) AmiGO: online access to ontology and annotation data. Bioinformatics 25(2): 288–289.
21. Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A (2010) Gene expression atlas at the european bioinformatics institute. Nucleic Acids Res 38(Database issue): D690–D698.
22. Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farne A, et al. (2012) Gene Expression Atlas update–a value-added database of microarray and sequencing-based functional genomics experiments. Nucleic Acids Res (Database issue): D1077–D1081.
23. Castro MA, Filho JL, Dalmolin RJ, Sinigaglia M, Moreira JC, et al. (2009) ViaComplex: software for landscape analysis of gene expression networks in genomic context. Bioinformatics 25(11): 1468–1469.
24. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4: 2.
25. Scardoni G, Petrerlini M, Laudanna C (2009) Analyzing biological network parameters with CentiScaPe. Bioinformatics 25(21): 2857–2859.
26. Newman MEJ (2005) A measure of betweenness centrality based on random walks. Soc Networks 27: 39–54.
27. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21(16): 3448–3449.
28. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57: 289–300.
29. Rosado JO, Henriques JP, Bonatto D (2011) A systems pharmacology analysis of major chemotherapy combination regimens used in gastric cancer treatment: predicting potential new protein targets and drugs. Curr Cancer Drug Targets 11(7): 849–869.
30. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3(4): e59.
31. van der Toorn M, Rezayat D, Kauffman HF, Bakker SJ, Gans RO, et al. (2009) Lipid-soluble components in cigarette smoke induce mitochondrial production of reactive oxygen species in lung epithelial cells. Am J Physiol Lung Cell Mol Physiol 297(1): L109–L114.
32. Menon R, Fortunato SJ, Yu J, Milne GL, Sanchez S, et al. (2011) Cigarette smoke induces oxidative stress and apoptosis in normal term fetal membranes. Placenta 32(4): 317–322.
33. Yao H, Yang SR, Kode A, Rajendrasozhan S, Caito S, et al. (2007) Redox regulation of lung inflammation: role of NADPH oxidase and NF-kappaB signaling. Biochem Soc Trans 35(Pt 5): 1151–1155.
34. Yao H, Edirisinghe I, Yang SR, Rajendrasozhan S, Kode A, et al. (2008) Genetic ablation of NADPH oxidase enhances susceptibility to cigarette smoke-induced lung inflammation and emphysema in mice. Am J Pathol 172(5): 1222–1237.
35. Haeggström JZ, Funk CD (2011) Lipoxygenase and leukotriene pathways: biochemistry, biology, and roles in disease. Chem Rev 111(10): 5866–5898.
36. Ylikorkala O, Viinikka L (1992) The role of prostaglandins in obstetrical disorders. Baillieres Clin Obstet Gynaecol 6(4): 809–827.
37. Lantz RC, Hays AM (2006) Role of oxidative stress in arsenic-induced toxicity. Drug Metab Rev 38(4): 791–804.
38. Board PG, Menon D (2012) Glutathione transferases, regulators of cellular metabolism and physiology. Biochim Biophys Acta.
39. Moledina F, Clarke G, Oskooei A, Onishi K, Günther A, et al. (2012) Predictive microfluidic control of regulatory ligand trajectories in individual pluripotent cells. Proc Natl Acad Sci U S A 109(9): 3264–3269.
40. Loureiro B, Oliveira LJ, Favoreto MG, Hansen PJ (2011) Colony-stimulating factor 2 inhibits induction of apoptosis in the bovine preimplantation embryo. Am J Reprod Immunol 65(6): 578–588.
41. Kim YJ, Lee GS, Hyun SH, Ka HH, Choi KC, et al. (2009) Uterine expression of epidermal growth factor family during the course of pregnancy in pigs. Reprod Domest Anim 44(5): 797–804.
42. Shilo BZ (2005) Regulating the dynamics of EGF receptor signaling in space and time. Development 132(18): 4017–4027.
43. Delpretti S, Zakany J, Duboule D (2012) A function for all posterior Hoxd genes during digit development? Dev Dyn 241(4): 792–802.
44. Singh AB, Harris RC (2005) Autocrine, paracrine and juxtacrine signaling by EGFR ligands. Cell Signal 17(10): 1183–1193.
45. Nyhan WL (2005) Disorders of purine and pyrimidine metabolism. Mol Genet Metab 86(1–2): 25–33.
46. Goggin M, Sangaraju D, Walker VE, Wickliffe J, Swenberg JA, et al. (2011) Persistence and repair of bifunctional DNA adducts in tissues of laboratory animals exposed to 1,3-butadiene by inhalation. Chem Res Toxicol 24(6): 809–817.
47. Koturbash I, Scherhag A, Sorrentino J, Sexton K, Bodnar W, et al. (2011) Epigenetic mechanisms of mouse interstrain variability in genotoxicity of the environmental toxicant 1,3-butadiene. Toxicol Sci 122(2): 448–456.
48. Kauffman SL (1969) Cell proliferation in embryonic mouse neural tube following urethane exposure. Dev Biol 20(2): 146–157.
49. Breiling A, Sessa L, Orlando V (2007) Biology of polycomb and trithorax group proteins. Int Rev Cytol 258: 83–136.

50. Matarazzo MR, De Bonis ML, Strazzullo M, Cerase A, Ferraro M, et al. (2007) Multiple binding of methyl-CpG and polycomb proteins in long-term gene silencing events. J Cell Physiol 210(3): 711–719.
51. Ma L, Zheng LW, Sham MH, Cheung LK (2010) Uncoupled angiogenesis and osteogenesis in nicotine-compromised bone healing. J Bone Miner Res 26(6): 1305–1313.
52. Lopatina N, Haskell JF, Andrews LG, Poole JC, Saldanha S, et al. (2002) Differential maintenance and de novo methylating activity by three DNA methyltransferases in aging and immortalized fibroblasts. J Cell Biochem 84(2): 324–334.
53. Xu Y, Zhang Y, Cardell LO (2010) Nicotine enhances murine airway contractile responses to kinin receptor agonists via activation of JNK- and PDE4-related intracellular pathways. Respir Res 11: 13.
54. León Y, Sanchez JA, Miner C, Ariza-McNaughton L, Represa JJ, et al. (1995) Developmental regulation of Fos-protein during proliferative growth of the otic vesicle and its relation to differentiation induced by retinoic acid. Dev Biol 167(1): 75–86.
55. Srinivas H, Juroske DM, Kalyankrishna S, Cody D, Price RE, et al. (2005) c-Jun N-terminal kinase contributes to aberrant retinoid signaling in lung cancer cells by phosphorylating and inducing proteasomal degradation of retinoic acid receptor alpha. Mol Cell Biol 25(3): 1054–1069.
56. Han BC, Xia HF, Sun J, Yang Y, Peng JP (2010) Retinoic acid-metabolizing enzyme cytochrome P450 26a1 (CYP26A1) is essential for implantation - functional study of its role in early pregnancy. J Cell Physiol 223(2): 471–479.
57. Feltes BC, de Faria Poloni J, Bonatto D (2011) The developmental aging and origins of health and disease hypotheses explained by different protein networks. Biogerontology 12(4): 293–308.
58. David P, Sabapathy K, Hoffmann O, Idarraga MH, Wagner EF (2002) JNK1 modulates osteoclastogenesis through both c-Jun phosphorylation-dependent and -independent mechanisms. J Cell Sci 115(22): 4317–4325.
59. Yildiz D (2004) Nicotine, its metabolism and an overview of its biological effects. Toxicon 43: 619–632.
60. Kawakita A, Sato K, Makino H, Ikegami H, Takayama S, et al. (2008) Nicotine Acts on Growth Plate Chondrocytes to Delay Skeletal Growth through the α7 Neuronal Nicotinic Acetylcholine Receptor. PLoS One 3(12): e3945.
61. Nakayama Y, Mezawa M, Araki S, Sasaki Y, Wang S, et al. (2009) Nicotine suppresses bone sialoprotein gene expression. J Periodont Res 44: 657–663.
62. Benowitz NL, Hukkanen J, Jacob P 3rd (2009) Nicotine chemistry, metabolism, kinetics and biomarkers. Handb Exp Pharmacol 192: 29–60.