have identified that amyloid β-precursor protein is an androgen-inducible gene that could be a novel prognostic and therapeutic target of PCa by screening genes in the vicinity of both ARBSs and acetylated histone H3 (AcH3) sites (Takayama *et al.*, 2009).

In terms of transcriptome, cDNA microarray has been extensively used for years as a convenient and cost-effective technique for analyzing gene expression signatures. cDNA microarray, however, is not designed to identify novel genes without probes, thus it has limitations to analyze the expression of noncoding RNAs, including short RNAs. As some microRNAs (miRNAs) and novel RNA transcripts have been reported to be relevant to PCa (Louro *et al.*, 2007; Shi *et al.*, 2007), an alternative high-throughput transcriptome technology will be also required to analyze noncoding RNA expression as well as coding RNA profiles.

In this study, we investigated AR actions across the entire human genome to elucidate the androgen-mediated transcriptional network by performing ChIP-chip and cap analysis gene expression (CAGE). We could determine novel AR-mediated genomic actions, in particular, those dependent on antisense (AS) or intergenic transcription start sites (TSSs) that have not been previously identified by any other methods. Thus, our integrated approach identify novel androgen target genes and provides potential diagnostic and therapeutic molecular targets of PCa, which can be applied to the clinical management of the advanced disease.

Results

Genome-wide analysis of androgen-activated promoters in prostate cancer cells

CAGE involves the preparation and sequencing of concatemers of 20-nucleotide (nt) DNA tags that were derived from the 5′-end of capped mRNA, and mapping the CAGE tags to the genome (Shiraki *et al.*, 2003). This method is cost-effective and enables high-throughput genome-wide analyses of TSSs and promoter usage. For preparing CAGE samples, we stimulated LNCaP cells with R1881 (10 nM), which were cultured in a hormone-depleted medium for 72 h. We extracted total RNA from cells stimulated for 6 and 24 h, and synthesized first-strand cDNA by reverse transcription. After cutting and amplifying the 20-nt CAGE tags derived from the 5′-ends of mRNAs, we sequenced the tag-ligated concatemers. We then mapped the sequenced CAGE tags to the human genome (total tags at 0 h were 1 929 416; at 6 h, 682 111; and at 24 h, 586 083). Regions containing multiple tags were referred to as tag clusters (TCs); these clusters were diversely distributed throughout the genome. By using Fisher's exact test, we identified TCs whose distributions were significantly regulated by androgen. The total number of androgen-dependent TCs obtained was 13 110 ($P<0.01$); 1572 TCs were obtained using a stringent threshold (1e-6.11 with Bonferroni correction). Among the androgen-regulated TCs, 78.9% were upregulated and the remaining was downregulated at 6 or 24 h after stimulation. In the promoter region of a representative androgen target gene *FKBP5* (Velasco *et al.*, 2004), we detected a time-dependent increase in tag numbers by androgen stimulation (21.2, 80.6 and 474.3 tags accumulated per million tags (t.p.m.) at 0, 6 and 24 h after R1881 treatment, respectively) (Figure 1a). Androgen-dependent increase in CAGE TCs was also shown in the promoter regions of PSA/KLK3 and TMPRSS2 (Supplementary Figure 1). The 5′-termini of RefSeq genes are often different from the TSSs obtained by CAGE analysis (Katayama *et al.*, 2007). For example, our study shows that TRIM36 includes an intronic CAGE TC within the second AcH3 site, which is situated in the downstream of the intronic ARBS (Supplementary Figure 2). Only 35% of CAGE TCs were located at a distance of <1 kb from the TSSs of known RefSeq genes (Figure 1b). In total, 13% of the TCs were identified at intergenic regions that were located >100 kb upstream of TSSs and did not overlap with any RefSeq gene. In all, 7% of the TCs were genome-widely distributed at the antisense regions of RefSeq genes, and 33% were transcribed from the RefSeq regions, most of which were from novel TSSs. Global transcriptome analysis using the CAGE technique and large-scale cDNA sequencing in the FANTOM3 project has recently shown that a large proportion of the mammalian genome can transcribe genes from both sense (S) and antisense strands, and that many of the unknown transcripts are noncoding RNAs (Carninci *et al.*, 2005). Taken together, these findings suggest that androgen-regulated transcripts are widely distributed across annotated genes and noncoding RNAs.

We also compared the androgen responsiveness of our CAGE TCs with RefSeq genes. Among protein-coding genes with an increasing number of CAGE TCs (>20 t.p.m. by 24 h) at their promoter regions (<0.1 kb from TSS), 51% of genes were also significantly upregulated (>1.5-fold) by 24-h treatment with R1881 (1 nM) in LNCaP cells based on the microarray data set GSE14028 (Yu *et al.*, 2010, $P=2.0e-24$). In contrast, 40% of protein-coding genes with a decreasing number of TCs were significantly downregulated (<0.8-fold) by R1881 in LNCaP cells (Yu *et al.*, 2010, $P=2.7e-5$). On the basis of another microarray data GSE7868 (Wang *et al.*, 2007), 44% of the RefSeq genes with an increasing number of CAGE TCs were upregulated in LNCaP cells treated with dihydrotestosterone (100 nM) for 16 h ($P=4.2e-28$) and 45% of the RefSeq genes with a decreasing number of TCs were downregulated by dihydrotestosterone ($P=1.7e-36$).

Genome-wide identification of ARBSs in prostate cancer cells

ChIP-chip analysis revealed 6366 ($P<1e-4$), or 2872 ($P<1e-5$) specific ARBSs in the genome of LNCaP cells. To compare our ChIP-chip data with previous studies, we found that 77 and 74% of our AR binding ($P<1e-5$) were involved in the ChIP-chip data (Wang *et al.*, 2009) and the ChIP-Seq data (Yu *et al.*, 2010), respectively. We also identified 25 945 AcH3 sites ($P<1e-4$) as

epigenetic markers that were associated with activated promoters (Bernstein *et al.*, 2005). Among 2872 ARBSs ($P<1e-5$), 420 ARBSs (15%) were overlapped with AcH3 sites. In addition, approximately 37.6% of AcH3 sites were mapped to the proximal regions of the TSSs of RefSeq genes (<3 kb from the TSSs), including many sites located in the downstream proximal regions of TSSs (Figure 1c), as consistent with a previous report in regard to genome-wide analysis of active promoters (Kim *et al.*, 2005). In terms of the overlap between the AcH3 sites and the CAGE TCs, 41% of the CAGE TCs were overlapped with AcH3 sites and 49% were overlapped with AcH3 sites within a distance <3 kb. On the other hand, most of the ARBSs were found to be widely distributed across the human genome, and only 4.8% of the ARBSs were mapped to promoter regions at <1 kb from the TSSs (Figure 1c).

We randomly selected 25 ARBSs in the vicinity of known androgen targets such as ACSL3, and intergenic regions on chromosomes 15, 20 and X to perform conventional ChIP analysis for AR binding in LNCaP cells (Figure 2a). When compared with four genomic regions as negative controls, most of the investigated regions were confirmed to be androgen-dependent ARBSs.

To further investigate the effects of the histone acetylation status on the AR-dependent transcriptional activity, we performed ChIP analysis for AR, SRC1, AcH3 and RNA Pol II recruitment for the group of ARBSs that overlapped with AcH3 sites (AcH3 overlap (+)) and the group of ARBSs distinct from the AcH3 sites (AcH3 overlap (−)) (Figure 2b). Significant androgen-dependent recruitment of both SRC1 and RNA Pol II was detected (>1.5-fold) in the AcH3 overlap (+) group, but not in the AcH3 overlap (−) group.

ChIP-chip and CAGE analyses reveal AR-dependent regulation of androgen target genes

Novel ARBSs and AcH3 sites adjacent to CAGE TCs were identified in the androgen-regulated genes. For instance, we found novel ARBSs in the intronic regions of α-methyl-CoA racemase, which is a prognostic marker for PCa (Luo *et al.*, 2002) (Figure 3a). Interestingly, AR binding was also observed in an intronic region of insulin-like growth factor-1 receptor, whose expression has been previously reported to be dependent on the nongenomic action of androgen (Pandini *et al.*,

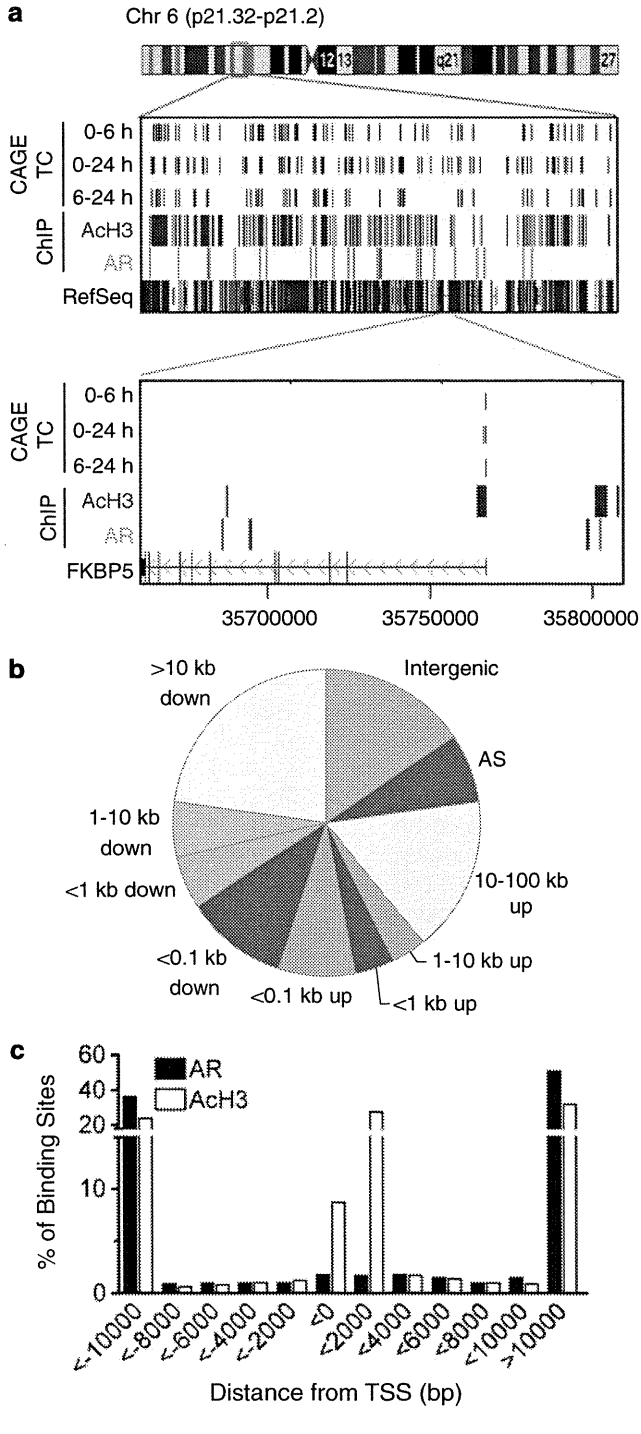


Figure 1 Genome-wide cap analysis of gene expression (CAGE) and chromatin immunoprecipitation on DNA chips (ChIP-chip) of the androgen-regulated gene network in hormone-naïve prostate cancer LNCaP cells. (**a**) Mapping of androgen-regulated promoters and androgen receptor binding sites (ARBSs) or acetylated histone H3 (AcH3) sites in the vicinity of FKBP5 on chromosome 6. Top panel, cytogenetic location of the representative androgen-regulated target gene FKBP5 on chromosome 6. Middle panel, UCSC genome browser view of androgen-dependent CAGE tag clusters (TCs) and AR/AcH3 sites determined by ChIP-chip at the 6p21 region. Androgen-upregulated CAGE TCs were mapped for three different time points after R1881 treatment (Tx) (0–6 h Tx, 6-h Tx versus 0-h; 0–24 h Tx, 24-h Tx versus 0-h; 6–24 h Tx, 24-h Tx versus 6-h). Bottom panel, browser view of CAGE TCs and AR/AcH3 sites in the vicinity of *FKBP5* gene and its promoter region. (**b**) Distribution of androgen-regulated CAGE TCs ($n=13{,}110$, $P<0.01$) relative to the transcription start sites (TSSs) of the RefSeq genes. TCs located >100 kb upstream of the TSSs of the RefSeq genes are classified as intergenic TCs. AS represents the TCs in the antisense regions of the RefSeq genes. (**c**) Distances of the AR/AcH3 sites from the TSSs of RefSeq genes. The *y* axis represents the percentages of the total sites that were identified as ARBSs or AcH3 sites.

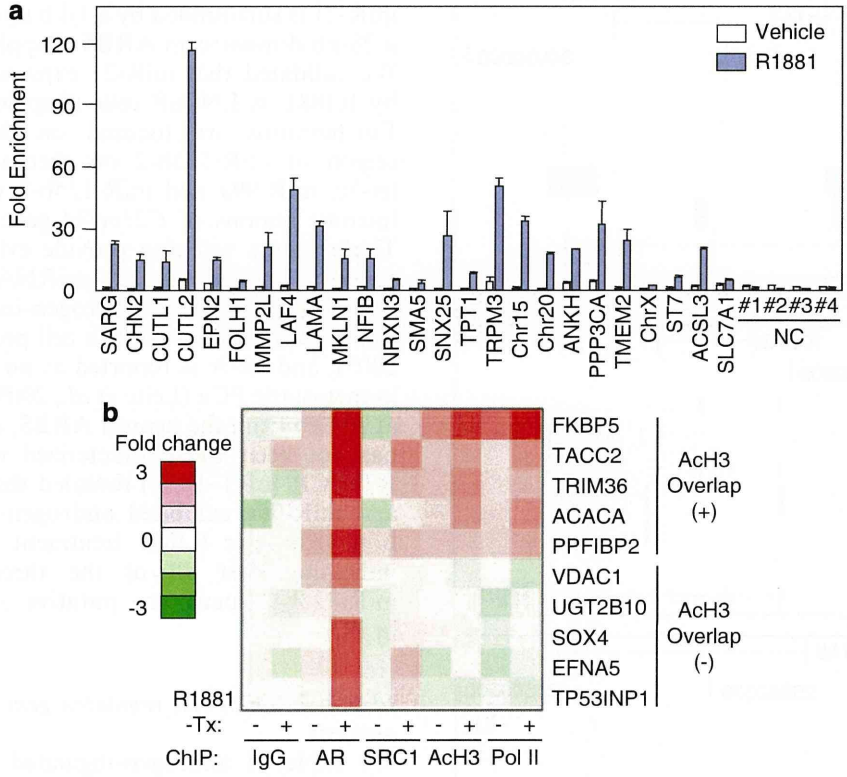


Figure 2 ChIP–qPCR validation of ARBSs determined by ChIP-chip. (**a**) Conventional ChIP analysis of 25 randomly selected ARBSs. AR binding was determined in LNCaP cells treated with R1881 or vehicle for 24 h by ChIP. NC represents negative control locus that did not show any specific AR binding (Horie-Inoue *et al*., 2006). ARBSs are denoted as the gene symbols of the RefSeq genes to which they were adjacent to as the respective chromosome numbers if they were located in the intergenic regions. (**b**) Effect of AcH3 sites overlap with ARBS on the binding of androgen-regulated transcriptional factors. Representative ARBSs with or without AcH3 sites overlap were examined for their occupancy of AR, SRC1, AcH3 and RNA Pol II in the promoter regions of the indicated genes. IgG was used as a control for nonspecific binding.

2009); thus, we presume that the genomic activity of AR will also contribute to the transcriptional modification of insulin-like growth factor-1 receptor (Figure 3b). Identification of ARBSs in other androgen-regulated genes, such as NK3 homeobox 1 (Prescott *et al*., 1998) (Figure 3c) and jagged 1 (Santagata *et al*., 2004) (Figure 3d), also provide further evidence that direct actions of AR is required for their transcriptional regulation.

S/AS transcriptional regulation by androgen showed diverse alteration patterns

Genome-wide promoter exchange such as bidirectional or sense/antisense (S/AS) transcriptional regulation is known to occur in the human genome. Antisense transcripts associate with neighboring genes in complex loci to form chains of linked transcriptional units (Katayama *et al*., 2005). We investigated the number of antisense-directed TCs that were associated with ARBSs: of the 1572 TCs, 173 (11% with Bonferroni correction) were situated on the antisense strand of RefSeq genes and 34 of these 173 TCs (19%) were associated with ARBSs within a 100-kb region ($P<1e\text{-}5$).

Antisense TCs were mostly abundant in the 3′-regions of RefSeq genes (Figure 4a). However, sense TCs were abundant in the proximal regions of the TSSs of RefSeq genes (Figure 4b). We identified 39 RefSeq genes with bidirectional S/AS pairs of CAGE TCs in the human genome. Among these S/AS pairs, 15 of 39 S/AS pairs exhibited reciprocal alteration of tag numbers in response to androgen, whereas other pairs exhibited concordant alteration. With regard to the androgen-dependent reciprocal patterns of S/AS pairs, one pattern consisted of an upregulated sense TC present in the TSS region of a RefSeq gene and a downregulated intronic antisense TC; this pattern was observed for the androgen-inducible gene *DNM1L* (Figure 4c). The other pattern consisted of a downregulated sense TC present in the TSS region of a RefSeq gene and an upregulated intronic sense TC, as exemplified by the androgen-repressed gene *SLC7A1* (Figure 4d). The contribution of an antisense promoter to the expression of a gene needs to be studied; nevertheless, CAGE analysis is a useful technology to clarify the complexity of androgen-regulated transcription in PCa cells.

ChIP-chip analysis reveals miRNA regulation by AR

Previous studies on expression profiling in PCa have revealed that ~20-nt short RNAs or miRNAs are closely associated with the progression of the disease.

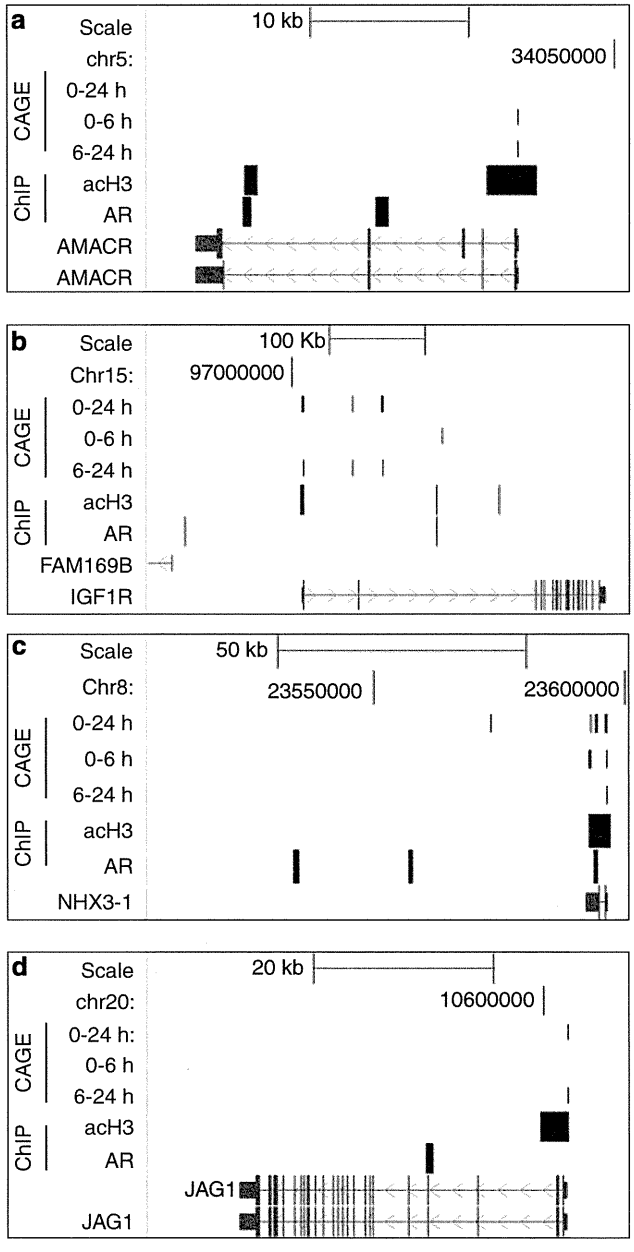


Figure 3 Contribution of AR-dependent promoters and AR occupancy to the transcriptional regulation of representative androgen-regulated genes. UCSC genome browser views of androgen-dependent CAGE TCs and AR/AcH3 sites determined by ChIP-chip in the vicinity of AMACR (**a**), IGF1R (**b**), NKX3.1 (**c**) and JAG1 (**d**).

On the basis of ChIP-chip data, we searched miRNAs in the vicinity of ARBSs; we identified four miRNAs in three miRNA clusters located at a distance of <100 kb from the ARBSs. We also identified four additional miRNAs that were relatively adjacent to the ARBSs or AcH3 sites, which were recently shown to be relevant in the progression of PCa (Shi *et al.*, 2007; Leite *et al.*, 2009; Ribas *et al.*, 2009; Sun *et al.*, 2009) (Table 1). Among them, miR-21 is a characterized AR-regulated miRNA that promotes hormone-dependent and -independent PCa growth (Ribas *et al.*, 2009). In our analysis, miR-21 is surrounded by a 1-kb upstream AcH3 site and a 28-kb downstream ARBS (Supplementary Figure 3a). We validated that miR-21 expression was upregulated by R1881 in LNCaP cells (Supplementary Figure 3b). Furthermore, we focused on the miRNA-clustered region of miR-125b-2 on chromosome 21; miRNAs let-7c, miR-99a and miR-125b-2 were clustered in the intronic regions of *C21orf34* gene (Figures 5a and b). These results will also provide evidence that androgen genomic actions regulate miRNAs, as miR-125b has been shown to be an androgen-inducible miRNA that can be associated with PCa cell proliferation (Shi *et al.*, 2007), and let-7c is reported as an upregulated miRNA in metastatic PCa (Leite *et al.*, 2009). miR-99a is located at 131 kb from the nearest ARBS, although the miRNA has not been well-characterized in PCa. Quantitative RT–PCR (qRT–PCR) revealed that miR-125b-2, let-7c and miR-99a exhibited androgen-dependent upregulation 48 h after R1881 treatment (Figure 5c), thereby indicating that all of the three miRNAs on the miR-125b-2 locus are putative AR-regulated targets in PCa cells.

Identification of AR regulated genes by ChIP-chip analysis

We explored androgen-regulated genes based on the data of ChIP-chip and CAGE analyses. Within a region of 100 kb from the ARBSs, 44, 66 and 134 androgen-upregulated CAGE TCs were identified by the comparison of time points at 6 versus 0 h, 24 versus 6 h and 24 versus 0 h, respectively (Figure 6a). mRNA microarray analysis revealed that 254 and 1477 genes were upregulated ($>$twofold) after 6 and 24 h of androgen treatment, respectively (Figure 6b). Among the genes upregulated by androgen, the rate of ARBS positivity (at <100 kb) was 21.6% at 24 h and 7.2% at 6 h (Figure 6b). AR binding bias was not significantly shown in upregulated genes at 6 h only versus the genomic background ($P=0.23$). In contrast, ARBS positivity was highly enriched in the upregulated genes at 24 h versus the genomic background ($P=1.4e\text{-}19$). These results show that the AR binding will contribute to the androgen-dependent transcription by 24 h. Moreover, representative qRT–PCR data showed that known and putative androgen target genes exhibited time-dependent upregulation and that the androgen-dependent upregulation was continuously observed up to at least for 48 h after treatment (Figure 6c). Taken together, we identified 258 novel upregulated genes ($>$twofold change at 24-h treatment), which may be putative direct targets of AR.

Androgen target genes are associated with tumorigenesis

To elucidate the role of putative androgen target genes in the tumorigenesis of PCa, we reviewed the expression profiles of 508 genes that were differentially regulated in our experiments (258 androgen-upregulated genes and 250 androgen-downregulated genes) in normal and cancer (PCa) regions of prostate tissues. cDNA microarray data for 62 primary tumors and 41 normal tissue

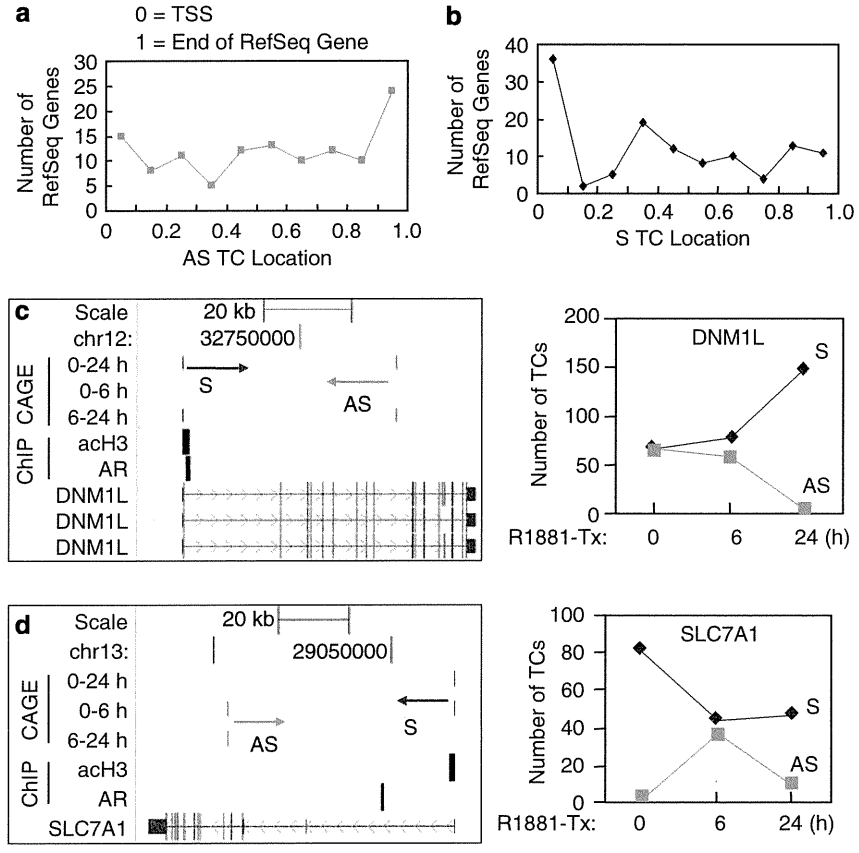


Figure 4 Identification of androgen-regulated CAGE TCs present in the antisense regions. (**a**) Distribution of antisense TCs relative to the proportional positions of RefSeq genes. Positions 0 and 1 on the *x* axis represent the TSSs and ends of RefSeq Genes, respectively. (**b**) Distribution of sense (S) TCs relative to the proportional positions of RefSeq genes. (**c**, **d**) Reciprocal regulation of S/AS promoter pairs of RefSeq genes in response to androgen. The sense promoter at the TSS of DNM1L was upregulated, whereas the antisense promoter at an intron of DNM1L was downregulated (**c**). The sense promoter at the TSS of SLC7A1 was downregulated, whereas the antisense promoter at an intron of SLC7A1 was upregulated (**d**).

Table 1 miRNAs adjacent to ARBSs

miRNA	*chr*	*Location*	*Region*	*ARBS (kb)*	*AcH3 (kb)*
miR-222	X	45362675–45362784	Intergenic	9.6	103.3
miR-221	X	45361839–45361948	Intergenic	10.5	104.2
miR-21	17	55273409–55273480	3′-downstream (TMEM49)	28.3	1.0
miR-125b-2	21	16884428–16884516	Intron (C21orf34)	80.2	1.1
miR-218-1	4	20206167–20206276	Intron (SLIT2)	111.3	172.2
miR-218-2	5	168127729–168127838	Intron (SLIT3)	116.9	151.7
let-7c	21	16834019–16834102	Intron (C21orf34)	130.6	84.7
miR-100	11	121528147–121528226	Intron (LOC399959)	307.2	42.3

Abbreviations: AcH3, histone H3 acetylated; ARBSs, androgen receptor binding sites; chr, chromosome; miRNA, microRNA.

samples (Lapointe *et al.*, 2004) were downloaded from the Stanford University website (http://microarray-pubs.stanford.edu/prostateCA); genes upregulated and downregulated by androgen were hierarchically clustered using the pairwise average-linkage method depending on the gene expression features of normal against tumor predominancy (Figure 6d). The cluster of genes with higher expression in PCa tissues than in normal prostate tissues is depicted as normal <PCa (red square), and the other cluster of genes with higher expression in normal prostate than in PCa samples is depicted as normal >PCa (blue square). The normal <PCa cluster containing androgen-upregulated genes (group I) included genes such as *ACACA* and *α-methyl-CoA racemase* (related to amino acid metabolism) and the normal >PCa cluster containing androgen-upregulated genes (group II) included genes such as *EGFR* and *EFNA5* (related to signal transduction). The normal <PCa cluster containing androgen-downregulated genes (group III) included genes such as *NFIX* and *SOX4* (related to transcriptional regulation), and the normal >PCa cluster containing androgen-downregu-

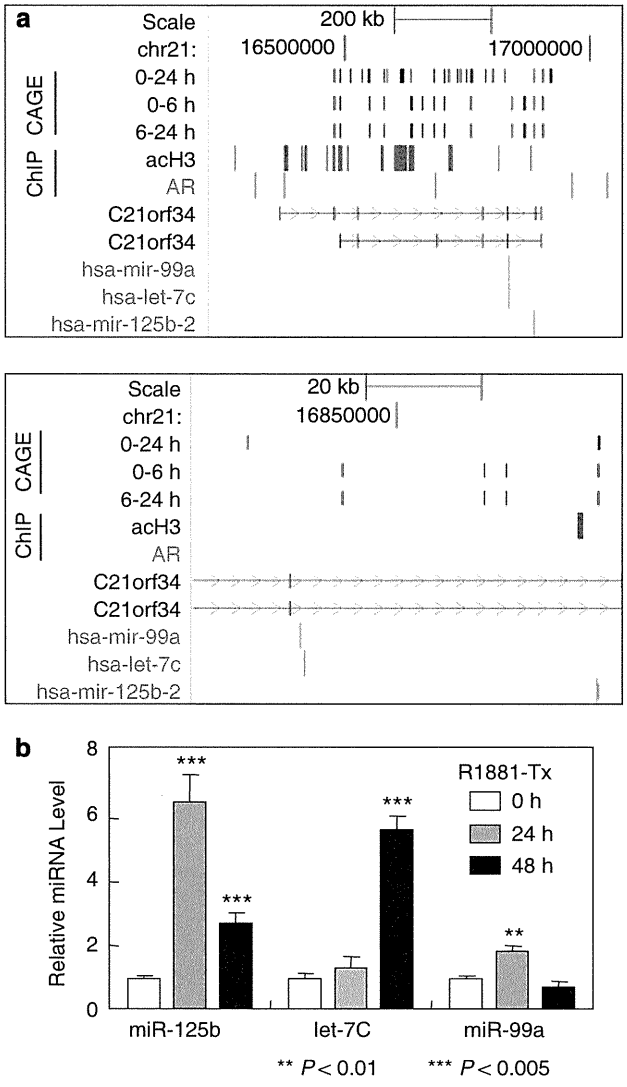


Figure 5 Androgen-dependent transcriptional regulation of miR-125b-2 microRNA locus on chromosome 21. (**a**) UCSC genome browser view of the miR-125b-2 locus, including hsa-mir-99a, hsa-let-7c and hsa-mir-125b-2. (**b**) Androgen-dependent expression of miRNAs in the miR-125b-2 locus. qRT–PCR was performed using TaqMan probes; the data represent relative miRNA levels normalized to the level of small nuclear RNA U6.

lated genes (group IV) included genes such as *TGFBR3* and *FOXO1A* (related to signal transduction and transcriptional regulation). Among the androgen-upregulated genes, the proportion of group I genes was significantly higher than that of group II genes. The proportion of group IV was also substantially higher than that of group III genes among the androgen-downregulated genes. The rates of the genes with the expressions altered by androgen were significantly higher in the clinical microarray data overlapped with LNCaP data compared with the randomly selected genes overlapped with the clinical data ($P=3.5e\text{-}9$ for upregulated genes and $P=0.001$ for downregulated genes). This clustering analysis of microarray data will indicate that the androgen-dependent transcriptional regulation is closely associated with the development of PCa.

Finally, we extracted novel androgen-inducible genes that are regulated by direct AR actions. We selected 569 RefSeq genes that exhibited androgen-inducible CAGE TCs 24 h after treatment at <0.1 kb from their TSSs. Among them, 51% were induced by >1.5-fold at 24 h after R1881 treatment, indicating that the majority of CAGE TCs function as *bonafide* androgen-regulated promoters. We analyzed novel AR target genes with CAGE TCs in the promoter regions or within the genes. Of the AR target genes, 31% (82 genes) included androgen-regulated CAGE TCs (at least 20 t.p.m.) (Figure 6e). In the Oncomine database (Wilson and Giguère, 2007), we identified 112 genes that were upregulated by androgen ($P<1e\text{-}4$) in PCa tissues (Figure 6f). Among these androgen-upregulated genes, 48 of 82 genes (58%) are associated with adjacent CAGE TCs, whereas 64 of 176 (36%) did not exhibit CAGE TCs ($P<0.001$).

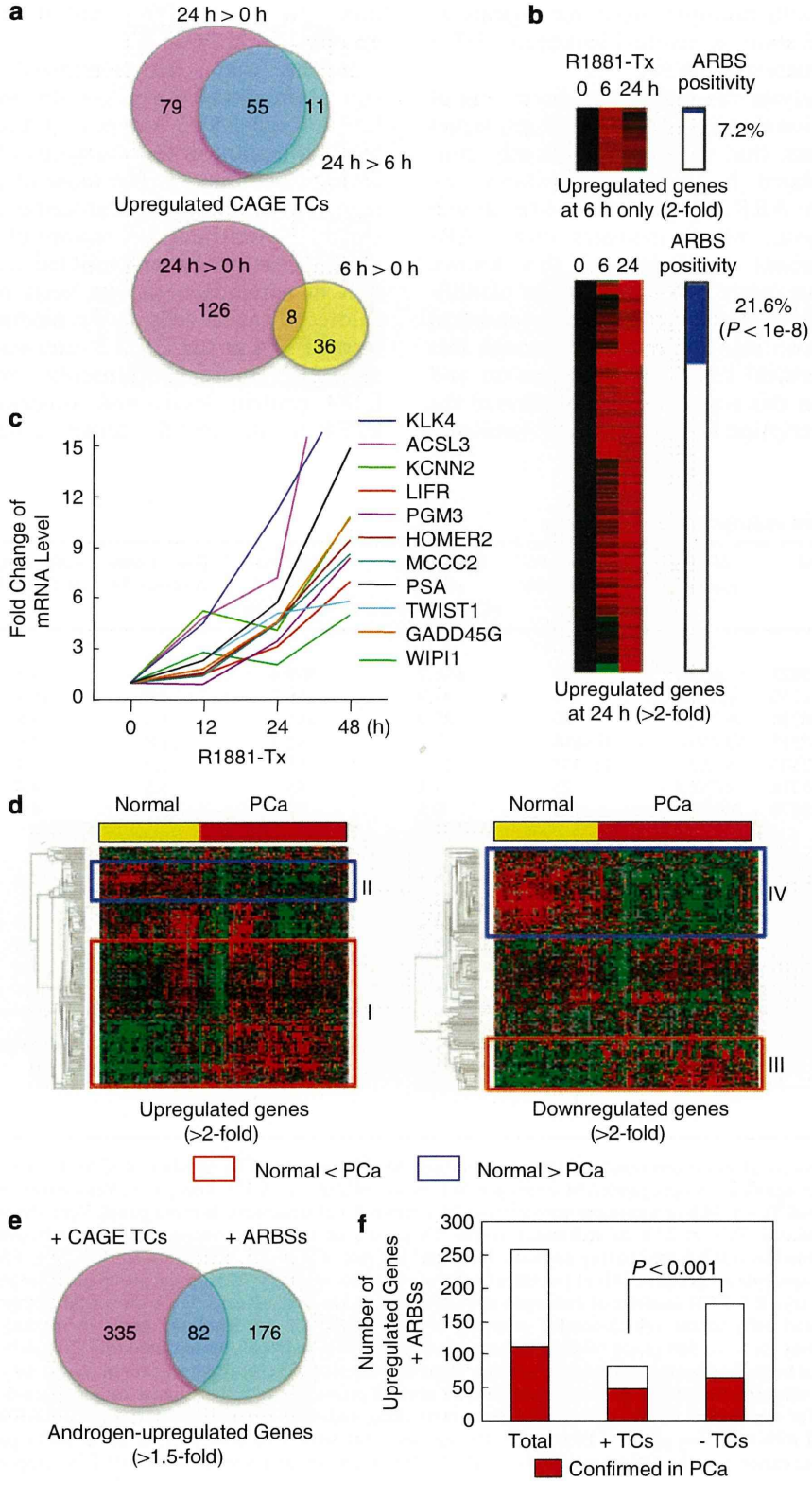
This study revealed a global approach for mapping of AR binding and AR-dependent promoters, which provides the systemic database regarding the AR-regulated gene network that essentially contributes to the development and progression of PCa. Notably, this study successfully confirmed that known androgen-regulated or PCa-specific genes are androgen-inducible, as exemplified by ACACA (Lapointe *et al.*, 2004), ABCC4 (Ho *et al.*, 2008), TRIM36 (Balint *et al.*, 2004) and amyloid β-precursor protein (Takayama *et al.*, 2009)(Table 2). On the basis of global mapping analysis, we identified novel androgen-upregulated genes, such as *FGFRL1*, *CAMKK2* and *C16orf60*, although their functions remain to be studied in PCa (Table 2). Notably, the application of pathway analysis using Database for Annotation, Visualization and Integrated Discovery Bioinformatics Resources 6.7 (Dennis *et al.*, 2003; Huang *et al.*, 2009) further shows that metabolic pathways, such as diabetes pathways, metabolism of lipids and lipoproteins, or amino sugar and nucleotide sugar metabolism, were enriched in our CAGE data (Supplementary Table 2).

Discussion

To understand the mechanism underlying AR regulation of the progression of PCa, we identified its transcriptional network in the genome of human PCa LNCaP cells. In the previous studies, we identified direct AR target genes by using the ChIP-cloning technique (Takayama *et al.*, 2008) or tiling arrays for the Encyclopedia Of DNA Elements regions of the human genome (Takayama *et al.*, 2007) or chromosomes 21 and 22 (Takayama *et al.*, 2009). A number of genes adjacent to ARBSs were found to be upregulated or downregulated by androgen. In this study, we used the high-throughput CAGE method for mapping androgen-regulated promoters in the human genome. To validate the present CAGE and ChIP-chip results, we mapped the data on the genome

browser adjacent to prototypic AR-regulated genes. Androgen treatment substantially increased the number of CAGE tags found in the promoter regions of AR-regulated genes such as *PSA/KLK3*, *TMPRSS2* and *FKBP5*.

We performed both CAGE and microarray analyses for comparing the two transcriptome technologies. Microarray analysis is useful for understanding global expression profiles of protein-coding genes, whereas CAGE is useful for detecting alterations in expressions



of individual promoters for both noncoding and protein-coding genes. The similarity between the two data sets is the expression alterations of protein-coding genes with CAGE tags in their promoters or within the genes. This tendency is well observed among the genes with CAGE tags in their promoters with a high TC. In contrast, the correlation between CAGE and microarray is lower for genes with multiple promoter regions as shown by the CAGE study in myeloid leukemia THP-1 cells (FANTOM Consortium, 2009).

Our combined analyses can clarify the direct role of AR in the transcriptional regulation of androgen target genes, including genes that were not previously considered to be regulated by androgen. Notably, we identified an intronic ARBS in the insulin-like growth factor-1 receptor gene, which indicates direct AR-mediated transcriptional regulation of this known androgen nongenomic target of androgen. The identification of a novel ARBS in the *JAG1* gene, a canonical ligand for Notch1, also suggests that AR through this binding site has a critical role in the progression and metastasis of PCa. In this study, we also confirmed the involvement of transcription factors in the AR-mediated gene network, as SOX4, one of the developmental transcription factors overexpressed in PCa (Liu *et al.*, 2006), was found to contain a functional ARBS within the gene. The calcium-dependent protein kinase CAMKK2 (Table 2) seems to be an interesting novel target of AR identified in this study. CAMKK2 can affect cell growth by upregulating AMPK phosphorylation (Yu *et al.*, 2009), and it is also known to be upregulated in PCa.

In this study, we determined that transcriptional regulation was also mediated by antisense promoters in LNCaP cells. The antisense-directed TSSs were relatively abundant in the 3′-regions of RefSeq genes. This finding is consistent with those of previous studies that reported that antisense transcripts are often transcribed from the 3′-untranslated regions of genes. The relevance of antisense promoter-mediated transcriptional regulation in carcinogenesis has been reported in HCT116 colorectal cancer cells, as the binding of β-Catenin to its target E2F4 at the E2F4 3′-untranslated region induces an E2F4 antisense transcript, which downregulates E2F4 protein levels and suppresses the binding of E2F4 to its specific target genes (Yochum *et al.*,

Table 2 Known and novel androgen-upregulated genes

Chr	*Start*	*End*	*Gene symbol*	*TC location (distance from TSS)*	*CAGE tag count at 0 h (t.p.m.)*	*CAGE tag count at 24 h (t.p.m.)*	*Fold change 6 versus 0 h*	*Fold change 24 versus 0 h*	*Function*
Known androgen targets									
21	26464837	26465023	*APP*	47	656.7	909.4	1.4	2.8	Cell proliferation
13	94751591	94751758	*ABCC4*	−4	39.9	158.7	2.9	11.1	Transport
17	32790079	32790210	*ACACA*	20	82.9	148.4	2.6	3.6	Metabolism
18	53967444	53967515	*NEDD4L*	104686	3.1	32.4	1.8	2.9	E3 ligase (HECT)
2	168812181	168812305	*STK39*	117337	20.2	63.1	1.5	3.4	Kinase
2	223551207	223551314	*ACSL3*	22	7.8	145	1.5	4.4	Metabolism
8	134378640	134378679	*NDRG1*	3	0.5	27.3	0.8	4.6	Differentiation
Novel androgen targets									
5	114533435	114533504	*TRIM36*	10638	8.3	42.7	2.1	2.6	E3 ligase (RING)
15	53369273	53369337	*RAB27A*	−10	8.8	68.2	2.3	6	Ras GTPase
1	36359091	36359198	*THRAP3*	−12	124.4	175.7	1.8	2.5	Transcription
16	79597576	79597706	*C16orf60*	1	19.7	264.5	1	2.9	Unknown
12	120197208	120197268	*CAMKK2*	1599	17.1	102.4	3.1	7.4	Ca-dependent kinase
4	995386	995471	*FGFRL1*	−29	35.8	63.1	1	2.1	Cell adhesion

Abbreviations: Ca, calcium; CAGE, cap analysis of gene expression; Chr, chromosome; TC, tag count; TSS, transcription start site.

Figure 6 Identification of androgen-regulated genes exhibiting AR binding. (**a**) The number of CAGE TCs adjacent to ARBSs (<100 kb) that were significantly upregulated by androgen in LNCaP cells ($P<1e\text{-}6.11$). Top panel: Venn diagrams for the number of androgen-upregulated TCs at 24 h of treatment versus 0 h or 24 h versus 6 h of treatment. Bottom panel: Venn diagrams for the number of androgen-upregulated TCs at 24 h of treatment versus 0 h or 6 h of treatment versus 0 h. (**b**) Androgen-upregulated genes (>twofold) determined by mRNA microarray analysis. Note that the rate of ARBS positivity (<100 kb of the TSS) was 21.6% in the group of androgen-upregulated genes at 24 h of treatment, whereas this rate was only 7.2% in the group of androgen-upregulated genes at 6 h of treatment. (**c**) qRT–PCR analysis of androgen-upregulated genes in LNCaP cells. Data show fold change of mRNA levels in R1881 (10 nM)-treated cells versus vehicle-treated controls. (**d**) Expression of AR-regulated genes in normal prostate or cancer specimens. Microarray data for 508 genes selected by our experiments (258 androgen-upregulated and 250 androgen-downregulated genes) were retrieved from the Stanford University website. Note that the rate of genes predominant in cancer was higher in androgen-upregulated genes, whereas the rate of genes predominant in normal prostate was rather higher in androgen-downregulated genes. (**e**) Venn diagrams for the number of AR-upregulated genes containing androgen-activated CAGE TCs or ARBSs. CAGE TCs (>20 t.p.m.) were located within RefSeq genes or their promoter regions in the sense direction. (**f**) The rate of androgen-upregulated genes confirmed in prostate cancers by the Oncomine database ($P<1e\text{-}4$) is higher in the presence of CAGE TCs compared with the absence of CAGE TCs.

2007). We also found novel pairs of S/AS promoters in individual RefSeq genes. Both reciprocal and concordant alterations in promoter transcription are known to occur in PCa. Taken together, the abundant expression of androgen-regulated antisense promoters in LNCaP cells will indicate the physiological significance of antisense mechanisms in the AR-mediated gene network in PCa.

This integrated study also identified short RNAs or miRNAs that are regulated by AR. We observed that CAGE TCs and ARBSs accumulate in the vicinity of the miR-125b-2 cluster on chromosome 21. All miRNAs in the miR-125b-2 locus were upregulated by androgen. This result is consistent with a recent study in which let-7c was found to be overexpressed in PCa tissues with high Gleason score (Leite *et al.*, 2009). The global mapping of AR-mediated gene regulation will further reveal novel short RNAs involved in the pathophysiology of PCa. Whether the adjacent ARBSs directly regulate the miRNA expression remains to be studied by performing such as chromosome conformation capture experiments.

Hierarchical cluster analysis of prostate tissue samples from a published microarray database and a comparative study of the Oncomine database revealed that the androgen-dependent overexpression of AR target genes and the increased activity of target gene promoters that is mediated by androgen are well-related to the signature of PCa. In contrast, a large cluster of androgen-downregulated genes is associated with a feature of normal prostate tissues.

It is also notable that partner genes involved in PCa-related chimeric transcripts were validated as direct AR targets in our study. We have previously shown that FOXP1 is a transcription factor that is also regulated by ARBSs (Takayama *et al.*, 2008). Like the well-known TMPRSS2–ETV1 transcript in PCa, the upregulation of FOXP1–ETV1 transcript was reported in PCa cells (Hermans *et al.*, 2008). Other androgen-regulated genes such as *ACSL3* (Attard *et al.*, 2008) and *NDRG1* (Pflueger *et al.*, 2009) were also identified as putative AR targets genes in this study (Table 2). In PCa, ACSL3 is fused to ETV1, whereas NDRG1 is fused to ERG. Further characterization will discover new partner genes for fusions from our database.

In summary, our integrated study is a powerful tool for elucidating the diversity of the AR-regulated transcriptional network, which alters the expression of various types of target genes throughout the genome, especially those present in the intergenic and antisense regions of RefSeq genes. The strand-specific transcriptome data combined with AR binding information will provide a basic pipeline for the screening of novel AR target genes and expand the global information available on the genome-wide data of direct AR-mediated transcriptional regulation; it will also offer new insights into unknown nongenomic regulations of genes and identify cancer-regulated biomarkers of unknown mechanisms. 'Oncomining' of the genome-wide landscape database will identify new classes of molecular targets that can be exploited for improving the diagnosis and treatment of PCa.

Materials and methods

Cell culture and reagents

Androgen-sensitive human PCa LNCaP cells (ATCC, Manassas, VA, USA) were maintained in RPMI medium supplemented with 10% fetal bovine serum, 50 U/ml penicillin and 50 μg/ml streptomycin. Before androgen treatment, cells were cultured in phenol red-free medium containing 5% charcoal-stripped fetal bovine serum for 48–72 h. Antibodies for AR, AcH3 and SRC1 and RNA Pol II were used for ChIP assay as described previously (Takayama *et al.*, 2007).

CAGE

CAGE libraries derived from total RNAs of LNCaP cells before (0 h) or after 6- or 24-h treatment with R1881 (10 nM) were generated as previously described (Maeda *et al.*, 2008). Briefly, first-strand cDNAs, which were transcribed to the 5′-end of capped RNAs, were attached to CAGE 'bar code' tags (AATAG for 0 h, ATTAT for 6 h and ATTGG for 24 h), and digested into 20-nt tags by *Mme*I (New England Biolabs, Ipswich, MA, USA). The CAGE tags were concatenated, ligated to sequencer-specific DNA adaptors and analyzed by 454 FLX Sequencing (Roche Diagnostics, Penzberg, Germany). The positions of the CAGE tags on the human genome (NCBI version 35) were determined by using Vmatch alignment tool (by Professor Stefan Kurtz, University of Hamburg, Germany), and 1 929 416, 682 111 and 586 083 CAGE tags were determined from 0-, 6- and 24-h libraries, respectively. CAGE tags are grouped into TCs in which the member tags map to the same strand of a chromosome and overlap by at least 1 bp. Fisher's exact test was performed to compare tag distributions between the 0-, 6-, and 24-h samples; 2×2 contingency matrix constructed tag numbers in a TC against tag number in the other TCs by a sample pair. We determined whether there were significant differences ($P<0.01$) among tag distributions for each TC at 6 and 24 h after androgen treatment (total number of TCs = 13 110). Also using the Bonferroni correction ($P<0.01/n$; where n is total TCs, 381 554), we identified 1572 androgen-regulated TCs. For calculating genomic locations of TCs relative to the closest Refseq genes or Refseq on the antisense strand, we defined the location of TC that has maximal frequency at the TSS regions. All data will be released from the website of the Genome Network Platform at National Institute of Genetics, Japan (http://genomenetwork.nig.ac.jp/download/dataset_e.html).

Microarray

Total RNAs were extracted from LNCaP cells treated with R1881 (10 nM) for 6 and 24 h, using ISOGEN reagent (Nippon Gene, Tokyo, Japan). Affymetrix (Santa clara, CA, USA) U133Plus2.0 expression microarrays were used. Data analysis was performed using the Affymetrix Microarray Suite software. For comparing arrays, normalization was performed using data from all probe sets.

ChIP-chip

ChIP-chip analysis was performed in biological triplicates by using antibodies for AR and AcH3 as previously described (Takayama *et al.*, 2007). Briefly, *in vitro* transcription was performed twice for amplification of ChIP materials. Amplified DNA was fragmented, labeled with biotin and hybridized to Human Tiling 1.0R microarrays (7 Chip Set) (Affymetrix). Data were analyzed as described previously. Using a stringent P-value cutoff 1e-5, 2800 AR ChIP-enriched regions were identified as significant ARBSs. Using a stringent P-value cutoff 1e-4, 25 954 AcH3 ChIP-enriched regions were identified as significant AcH3 sites. For comparison study of our data with previous AR binding data, we retrieved the ChIP-chip

(Wang *et al.*, 2009) and ChIP-Seq (Yu *et al.*, 2010) data and mapped all the ARBS data to the human genome hg17 using liftOver tool in the UCSC Genome Bioinformatics (http://genome.ucsc.edu/).

ChIP and quantitative PCR (qPCR)
ChIP was performed as previously described (Takayama *et al.*, 2007). Fold enrichments relative to immunoglobulin G or the input control were quantified by real-time PCR using SYBR green PCR master mix (Applied Biosystems, Foster City, CA, USA) and the ABI Prism 7000 system (Applied Biosystems) based on SYBR green É fluorescence. Relative differences in the amounts of PCR products among the treatment groups were evaluated by the comparative cycle threshold (C_t) method, using glyceraldehyde-3-phosphate dehydrogenase as an internal control. Primer sequences for ARBSs are listed in the Supplementary Table 1.

Quantitative reverse transcription–PCR (qRT–PCR)
First-strand cDNA was synthesized using the Primescript RT reagent kit (TAKARA, Kyoto, Japan). Primer sequences are listed in the Supplementary Table 1. qRT–PCR for miRNA was performed in triplicates by TaqMan miRNA RT–PCR system (Ambion, Austin, TX, USA), using RUN6B (Applied Biosystems) as an internal control.

Hierarchical cluster analysis for androgen-regulated genes in prostate samples
Unsupervised hierarchical clustering was performed for 508 genes selected by our experiments (258 androgen-upregulated and 250 androgen-downregulated genes). cDNA microarray data for 62 primary tumors and 41 normal samples (Lapointe *et al.*, 2004) were retrieved from the Stanford University website (http://microarray-pubs.stanford.edu/prostateCA) and both androgen-upregulated and downregulated genes were clustered in a manner dependent on the gene expression features of normal or tumor predominancy, using a hierarchiral clustering by Cluster3.0 software (developed by Eisen *et al.*, 1998, extended by de Hoon *et al.*, 2004), based on the pairwise average-linkage method.

Pathway analysis
Pathway analysis was performed by Database for Annotation, Visualization and Integrated Discovery Bioinformatics Resources 6.7 (Dennis *et al.*, 2003; Huang *et al.*, 2009), using REACTOME (http://www.reactome.org/) and Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) databases as annotation sources.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This study was supported by the Genome Network Project, Cell Innovation Program and Support Project of Strategic Research Center in Private Universities from the MEXT, grants from the Japan Society for the Promotion of Science, grants-in-aid from the MHLW and the Program for Promotion of Fundamental Studies in Health Sciences of the NIBIO. We are grateful to T Murata for helpful discussion and Hiromi Sano and Kazumi Yamaguchi for their assistance.

References

Attard G, Clark J, Ambroisine L, Mills IG, Fisher G, Flohr P *et al.* (2008). Heterogeneity and clinical significance of ETV1 translocations in human prostate cancer. *Br J Cancer* **99**: 314–320.

Balint I, Muller A, Nagy A, Kovacs G. (2004). Cloning and characterisation of the RBCC728/TRIM36 zinc-binding protein from the tumor suppressor gene region at chromosome 5q22.3. *Gene* **332**: 45–50.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ *et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**: 169–181.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N *et al.* (2005). The transcriptional landscape of the mammalian genome. *Science* **309**: 1559–1563.

Chen CD, Welsbie DS, Tran C, Baek SH, Chen R, Vessella R *et al.* (2004). Molecular determinants of resistance to antiandrogen therapy. *Nat Med* **10**: 33–39.

de Hoon MJ, Imoto S, Nolan J, Miyano S. (2004). Open source clustering software. *Bioinformatics* **20**: 1453–1454.

Debes JD, Tindall DJ. (2004). Mechanisms of androgen-refractory prostate cancer. *N Engl J Med* **351**: 1488–1490.

Dehm SM, Tindall DJ. (2006). Molecular regulation of androgen action in prostate cancer. *J Cell Biochem* **99**: 333–344.

Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC *et al.* (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**: R3.

Eisen MB, Spellman PT, Brown PO, Botstein D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868.

FANTOM Consortium (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* **41**: 553–562.

Hermans KG, van der Korput HA, van Marion R, van de Wijngaart DJ, Ziel-van der Made A, Dits NF *et al.* (2008). Truncated ETV1, fused to novel tissue-specific genes, and full-length ETV1 in prostate cancer. *Cancer Res* **68**: 7541–7549.

Ho LL, Kench JG, Handelsman DJ, Scheffer GL, Stricker PD, Grygiel JG *et al.* (2008). Androgen regulation of multidrug resistance-associated protein 4 (MRP4/ABCC4) in prostate cancer. *Prostate* **68**: 1421–1429.

Horie-Inoue K, Takayama K, Bono HU, Ouchi Y, Okazaki Y, Inoue S. (2006). Identification of novel steroid target genes through the combination of bioinformatics and functional analysis of hormone response elements. *Biochem Biophys Res Commun* **339**: 99–106.

Huang DW, Sherman BT, Lempicki RA. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat Protoc* **4**: 44–57.

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M *et al.* (2005). Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564–1566.

Katayama S, Kanamori-Katayama M, Yamaguchi K, Carninci P, Hayashizaki Y. (2007). CAGE-TSSchip: promoter-based expression profiling using the 5′-leading label of capped transcripts. *Genome Biol* **8**: R42.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA *et al.* (2005). A high-resolution map of active promoters in the human genome. *Nature* **436**: 876–880.

Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K *et al.* (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci USA* **101**: 811–816.

Leite KR, Sousa-Canavez JM, Reis ST, Tomiyama AH, Camara-Lopes LH, Sanudo A *et al.* (2009). Change in expression of miR-let7c, miR-100, and miR-218 from high grade localized prostate cancer to metastasis. *Urol Oncol* (e-pub ahead of print).

Liu P, Ramachandran S, Ali Seyed M, Scharer CD, Laycock N, Dalton WB *et al.* (2006). Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer Res* **66**: 4011–4019.

Louro R, Nakaya HI, Amaral PP, Festa F, Sogayar MC, da Silva AM *et al.* (2007). Androgen responsive intronic non-coding RNAs. *BMC Biol* **5**: 4.

Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ *et al.* (2002). Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. *Cancer Res* **62**: 2220–2226.

Maeda N, Nishiyori H, Nakamura M, Kawazu C, Murata M, Sano H *et al.* (2008). Development of a DNA barcode tagging method for monitoring dynamic changes in gene expression by using an ultra high-throughput sequencer. *Biotechniques* **45**: 95–97.

Massie CE, Adryan B, Barbosa-Morais NL, Lynch AG, Tran MG, Neal DE *et al.* (2007). New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep* **8**: 871–878.

Pandini G, Genua M, Frasca F, Vigneri R, Belfiore A. (2009). Sex steroids upregulate the IGF-1R in prostate cancer cells through a nongenotropic pathway. *Ann NY Acad Sci* **1155**: 263–267.

Pflueger D, Rickman DS, Sboner A, Perner S, LaFargue CJ, Svensson MA *et al.* (2009). N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer. *Neoplasia* **11**: 804–811.

Prescott JL, Blok L, Tindall DJ. (1998). Isolation and androgen regulation of the human homeobox cDNA, NKX3.1. *Prostate* **35**: 71–80.

Ribas J, Ni X, Haffner M, Wentzel EA, Salmasi AH, Chowdhury WH *et al.* (2009). miR-21: an androgen receptor-regulated microRNA that promotes hormone-dependent and hormone-independent prostate cancer growth. *Cancer Res* **69**: 7165–7169.

Santagata S, Demichelis F, Riva A, Varambally S, Hofer MD, Kutok JL *et al.* (2004). JAGGED1 expression is associated with prostate cancer metastasis and recurrence. *Cancer Res* **64**: 6854–6857.

Shang Y, Myers M, Brown M. (2002). Formation of the androgen receptor transcription complex. *Mol Cell* **9**: 601–610.

Shi XB, Xue L, Yang J, Ma AH, Zhao J, Xu M *et al.* (2007). An androgen-regulated miRNA suppresses Bak1 expression and induces androgen-independent growth of prostate cancer cells. *Proc Natl Acad Sci USA* **104**: 19983–19988.

Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H *et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA* **100**: 15776–15781.

Sun T, Wang Q, Balk S, Brown M, Lee GS, Kantoff P. (2009). The role of microRNA-221 and microRNA-222 in androgen-independent prostate cancer cell lines. *Cancer Res* **69**: 3356–3363.

Suzuki H, Ueda T, Ichikawa T, Ito H. (2003). Androgen receptor involvement in the progression of prostate cancer. *Endocr Relat Cancer* **10**: 209–216.

Takayama K, Horie-Inoue K, Ikeda K, Urano T, Murakami K, Hayashizaki Y *et al.* (2008). FOXP1 is an androgen-responsive transcription factor that negatively regulates androgen receptor signaling in prostate cancer cells. *Biochem Biophys Res Commun* **374**: 388–393.

Takayama K, Kaneshiro K, Tsutsumi S, Horie-Inoue K, Ikeda K, Urano T *et al.* (2007). Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis. *Oncogene* **26**: 4453–4463.

Takayama K, Tsutsumi S, Suzuki T, Horie-Inoue K, Ikeda K, Kaneshiro K *et al.* (2009). Amyloid precursor protein is a primary androgen target gene that promotes prostate cancer growth. *Cancer Res* **69**: 137–142.

Velasco AM, Gillis KA, Li Y, Brown EL, Sadler TM, Achilleos M *et al.* (2004). Identification and validation of novel androgen-regulated genes in prostate cancer. *Endocrinology* **145**: 3913–3924.

Wang Q, Carroll JS, Brown M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell* **19**: 631–642.

Wang Q, Li W, Liu XS, Carroll JS, Janne OA, Keeton EK *et al.* (2007). A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell* **27**: 380–392.

Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J *et al.* (2009). Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**: 245–256.

Wilson BJ, Giguère V. (2007). Identification of novel pathway partners of p68 and p72 RNA helicases through Oncomine meta-analysis. *BMC Genomics* **8**: 419.

Yochum GS, Cleland R, McWeeney S, Goodman RH. (2007). An antisense transcript induced by Wnt/beta-catenin signaling decreases E2F4. *J Biol Chem* **282**: 871–878.

Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X *et al.* (2010). An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**: 443–454.

Yu SY, Chan DW, Liu VW, Ngan HY. (2009). Inhibition of cervical cancer cell growth through activation of upstream kinases of AMP-activated protein kinase. *Tumour Biol* **30**: 80–85.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)

Bousquet *et al. Genome Medicine* 2011, **3**:43
http://genomemedicine.com/content/3/7/43

Genome Medicine

CORRESPONDENCE Open Access

Systems medicine and integrated care to combat chronic noncommunicable diseases

Jean Bousquet[1]*, Josep M Anto[2], Peter J Sterk[3], Ian M Adcock[4], Kian Fan Chung[5], Josep Roca[6], Alvar Agusti[6], Chris Brightling[7], Anne Cambon-Thomsen[8], Alfredo Cesario[9], Sonia Abdelhak[10], Stylianos E Antonarakis[11], Antoine Avignon[12], Andrea Ballabio[13], Eugenio Baraldi[14], Alexander Baranov[15], Thomas Bieber[16], Joël Bockaert[17], Samir Brahmachari[18], Christian Brambilla[19], Jacques Bringer[20], Michel Dauzat[21], Ingemar Ernberg[22], Leonardo Fabbri[23], Philippe Froguel[24], David Galas[25], Takashi Gojobori[26], Peter Hunter[27], Christian Jorgensen[28], Francine Kauffmann[29], Philippe Kourilsky[30], Marek L Kowalski[31], Doron Lancet[32], Claude Le Pen[33], Jacques Mallet[34], Bongani Mayosi[35], Jacques Mercier[36], Andres Metspalu[37], Joseph H Nadeau[25], Grégory Ninot[38], Denis Noble[39], Mehmet Öztürk[40], Susanna Palkonen[41], Christian Préfaut[36], Klaus Rabe[42], Eric Renard[20], Richard G Roberts[43], Boleslav Samolinski[44], Holger J Schünemann[45], Hans-Uwe Simon[46], Marcelo Bento Soares[47], Giulio Superti-Furga[48], Jesper Tegner[49], Sergio Verjovski-Almeida[50], Peter Wellstead[51], Olaf Wolkenhauer[52], Emiel Wouters[53], Rudi Balling[54], Anthony J Brookes[55], Dominique Charron[56], Christophe Pison[57,58], Zhu Chen[59], Leroy Hood[25] and Charles Auffray[56,57,58,60,61]

Abstract

We propose an innovative, integrated, cost-effective health system to combat major non-communicable diseases (NCDs), including cardiovascular, chronic respiratory, metabolic, rheumatologic and neurologic disorders and cancers, which together are the predominant health problem of the 21st century. This proposed holistic strategy involves comprehensive patient-centered integrated care and multi-scale, multi-modal and multi-level systems approaches to tackle NCDs as a common group of diseases. Rather than studying each disease individually, it will take into account their intertwined gene-environment, socio-economic interactions and co-morbidities that lead to individual-specific complex phenotypes. It will implement a road map for predictive, preventive, personalized and participatory (P4) medicine based on a robust and extensive knowledge management infrastructure that contains individual patient information. It will be supported by strategic partnerships involving all stakeholders, including general practitioners associated with patient-centered care. This systems medicine strategy, which will take a holistic approach to disease, is designed to allow the results to be used globally, taking into account the needs and specificities of local economies and health systems.

*Correspondence: jean.bousquet@inserm.fr
[1]Department of Respiratory Diseases, Arnaud de Villeneuve Hospital, CHU Montpellier, INSERM CESP U1018, Villejuif, France
Full list of author information is available at the end of the article

Non-communicable diseases, the major global health problem of the century

Chronic diseases are disorders of long duration and generally slow progression [1]. They include four major non-communicable diseases (NCDs) listed by the World Health Organization (WHO) [2] – cardiovascular diseases, cancer, chronic respiratory diseases and diabetes – as well as other NCDs, such as neuropsychiatric disorders [3] and arthritis. As survival rates have improved for infectious and genetic diseases, chronic diseases have come to include communicable diseases (such as HIV/AIDS) and genetic disorders (such as cystic fibrosis). NCDs represent the major global health problem of the 21st century [4,5]; they affect all age groups [6] and their burden is greater than that of infectious diseases. NCDs are the world leading cause of disease burden and mortality [2] and are increasing in prevalence and burden [7], even in low- and middle-income countries [8]. Costs incurred by uncontrolled NCDs are substantial, especially in underserved populations [9] and low- and middle-income countries [10,11]. NCDs are an under-appreciated cause of poverty and hinder economic development [11]. Importantly, management of NCDs has recently been prioritized globally (Box 1).

Chronic diseases are caused by complex gene-environment interactions acting across the lifespan from the fetus to old age (Figure 1). In this context, 'environment' includes risk and protective factors associated with environment and lifestyle, such as

Box 1: Priorities for the prevention and control of NCDs

May 2008: 61st World Health Assembly. WHO recommended a worldwide priority policy on NCD prevention and control (2008 to 2013), including cardiovascular disease, cancer, chronic respiratory diseases [101] and diabetes, not least because they often have common environmental risk factors [2].

May 2010: United Nations (UN) General Assembly unanimously adopted Resolution A/RES/64/265: 'Tackling NCDs constitutes one of the major challenges for sustainable development in the 21st century' [102].

December 2010: the Council of the European Union adopted conclusions based on innovative and global approaches for NCDs in public health and healthcare systems to further develop population-based and patient-centered policies [1].

2010: US Center for Disease Control and Prevention (CDC) [103] says that 'an essential strategy for keeping older adults healthy is preventing NCDs and reducing associated complications'.

19 September 2011: UN General Assembly symposium on NCDs.

tobacco, nutrition, indoor and outdoor air pollution and sedentary life [2].

Socio-economic determinants are intertwined with the onset, progression, severity and control of NCDs. There are functional interdependencies between molecular components, reflecting complex network perturbations that link cells, tissues and organs [12]. Early life events are crucial in the generation of NCDs, and aging increases disease complexity, adding, for example, tissue and cell senescence [13]. Comorbidity refers to the co-existence of two or more diseases or conditions in the same individual that have similar risk factors and/or mechanisms. Most people with NCDs suffer from two or more diseases [14]. Co-morbidity and multi-morbidity are common signatures of NCDs and are associated with worse health outcomes [15], complex pharmacological interventions and clinical management, and increased healthcare costs [16]. However, little is known about how NCDs truly cluster at the genetic, molecular or mechanistic levels, and there is scant understanding of

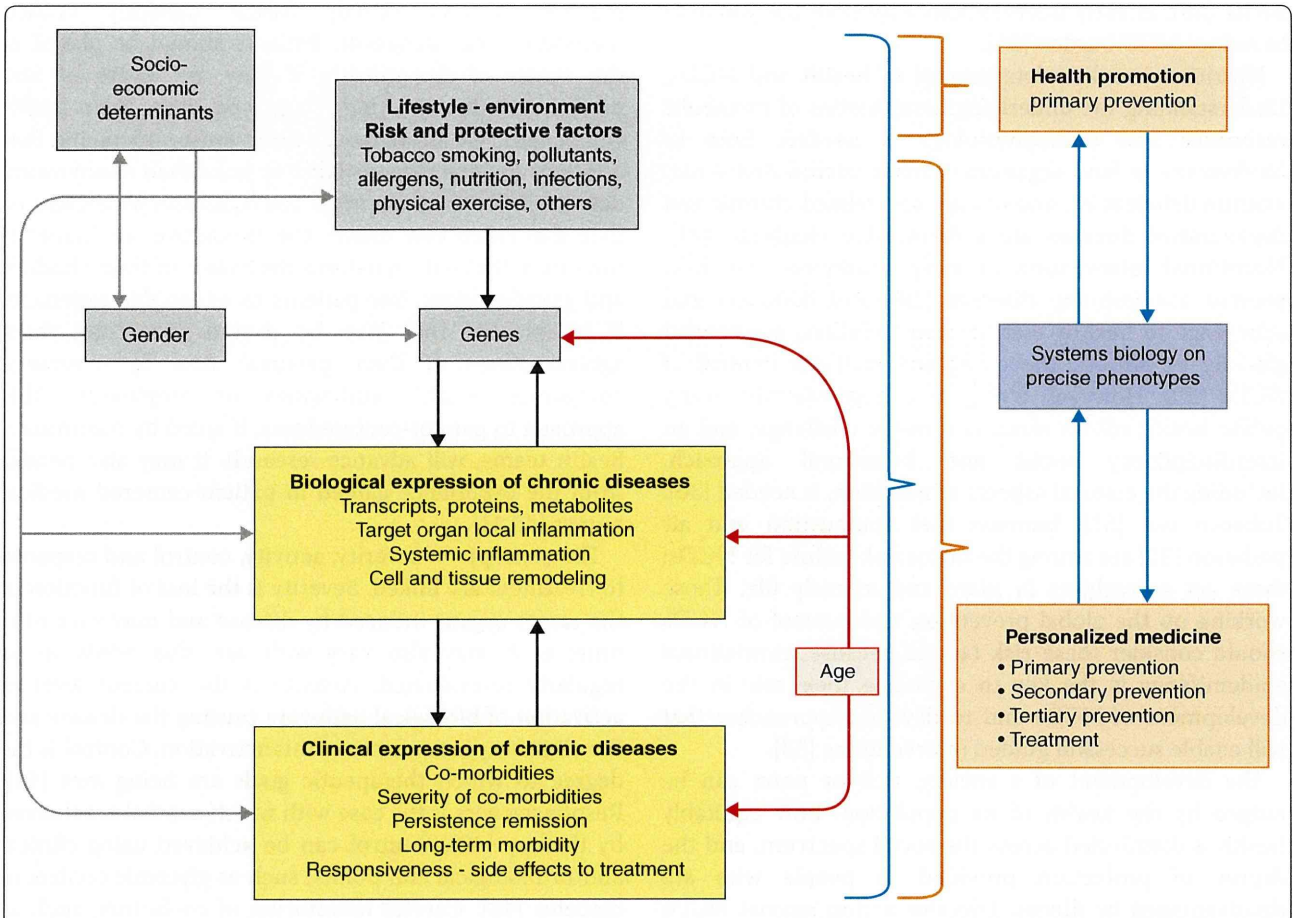


Figure 1. NCDs are associated with complex gene-environment interactions modulated by socio-economic determinants, psychological factors, age and gender. The products of these interactions lead to the biological expression of NCDs and further to their clinical expression with co-morbidities. A new definition of NCD phenotypes is needed to understand how a network of molecular and environmental factors can lead to complex clinical outcomes of NCDs for prevention and control.

how specific combinations of NCDs influence prognosis and treatment [16].

NCDs are multi-factorial. In addition to environmental factors and increased life expectancy, intrinsic host responses, such as local and systemic inflammation, immune responses and remodeling [17], have key roles in the initiation and persistence of diseases and co-morbidities. The recent increase in NCDs has been associated in part with biodiversity loss [18], socio-economic inequities linked with climate change, and loss of natural environments [19]. A more comprehensive understanding of these links will make it possible to propose more effective primary prevention strategies. The *in utero* environment is an important determinant of adult NCDs, including diabetes [20], coronary heart diseases [21], and asthma [22] or chronic obstructive pulmonary disease (COPD) [23]. Mechanistic links have been proposed that involve fetal expression of genes that are conserved across species, epigenetic mechanisms [22,24], early and maternal life infections, and/or environmental exposures. These need to be understood better [25], as early interventions may have the potential to reduce NCD burden [26].

Nutrition is a key determinant of health and NCDs. Understanding the underlying complexities of metabolic responses and pathophysiology is needed. Loss of biodiversity in food organisms causes micronutrient and vitamin deficiencies, and obesity and related chronic and degenerative diseases are a formidable challenge [27]. Nutritional intervention in early childhood may help prevent autoimmune diseases [28], and adoption and adherence to healthy diet recommendations are needed globally to prevent the onset and facilitate control of NCDs [29]. However, trying to change lifestyles using public health efforts remains a major challenge, and an interdisciplinary social and behavioral approach, including the cultural aspects of nutrition, is needed [30]. Tobacco use [31], biomass fuel combustion and air pollution [32] are among the major risk factors for NCDs; these act as early as *in utero* and in early life. Those working on the global prevention and control of NCDs should consider these risk factors because translational epidemiology is the key to exploring their role in the development of NCDs and to devising approaches that will enable successful guided interventions [33].

The development of a society, rich or poor, can be judged by the health of its population, how equitably health is distributed across the social spectrum, and the degree of protection provided to people who are disadvantaged by illness. Effective action against NCDs needs to include understanding of the social and economic determinants and their modification (Figure 1) [34]. Indeed, best-practice interventions targeted at coronary risk factors eliminate most socioeconomic differences that affect coronary heart disease mortality, and this should serve as an example to follow for other NCDs [35]. In May 2009, the 62nd WHO Assembly recommended re-orienting health systems globally to promote primary healthcare as the most cost-effective strategy [36]. Healthcare often focuses on single diseases, advanced technology, biomedical interventions and specialist care. Most healthcare takes place in primary care settings [37], with emphasis on providing a complete range of care, from home to hospital, and on investing resources rationally. Fragmenting care can reduce the ability of primary care clinicians to ensure that patient care is comprehensive, integrated, holistic, and coordinated [38], and to decide whether a person has a significant disease or temporary symptoms [39].

A proposal for multidisciplinary patient-centered management of chronic NCDs

We recommend that, to determine measures of disease severity and control, effective interventions and studies should be built around carefully phenotyped patients (Figure 2) and strictly follow carefully crafted methodological standards. Patients should be placed at the center of the system; if they are aware of and understand the resulting phenotype data, their health will benefit. We stress that patients must understand that it is their societal responsibility to make their anonymized data available to appropriate scientists and physicians so that the latter can create the predictive medicine of the future that will transform the health of their children and grandchildren. For patients to adopt this approach, it is essential that laws be passed protecting them against abuse of their personal data by insurance companies, health authorities or employers. This approach to patient-centeredness, if aided by community health teams, will advance research. It may also benefit from the experience gained in patient-centered medical homes [40,41].

The concepts of severity, activity, control and response to treatment are linked. Severity is the loss of function in the target organs induced by disease and may vary over time; as it may also vary with age, this needs to be regularly re-evaluated. Activity is the current level of activation of biological pathways causing the disease and the clinical consequences of this activation. Control is the degree to which therapeutic goals are being met [42]. Responsiveness is the ease with which control is achieved by therapy [43]. Control can be achieved using clinical and/or biological end points, such as glycemic control in diabetes [44]. Careful monitoring of co-factors, such as compliance, and of unavoidable risk factors is needed. The uniform definition of severe asthma presented to WHO is based on this approach [45] and therefore provides a model to assess NCD severity (Figure 3).

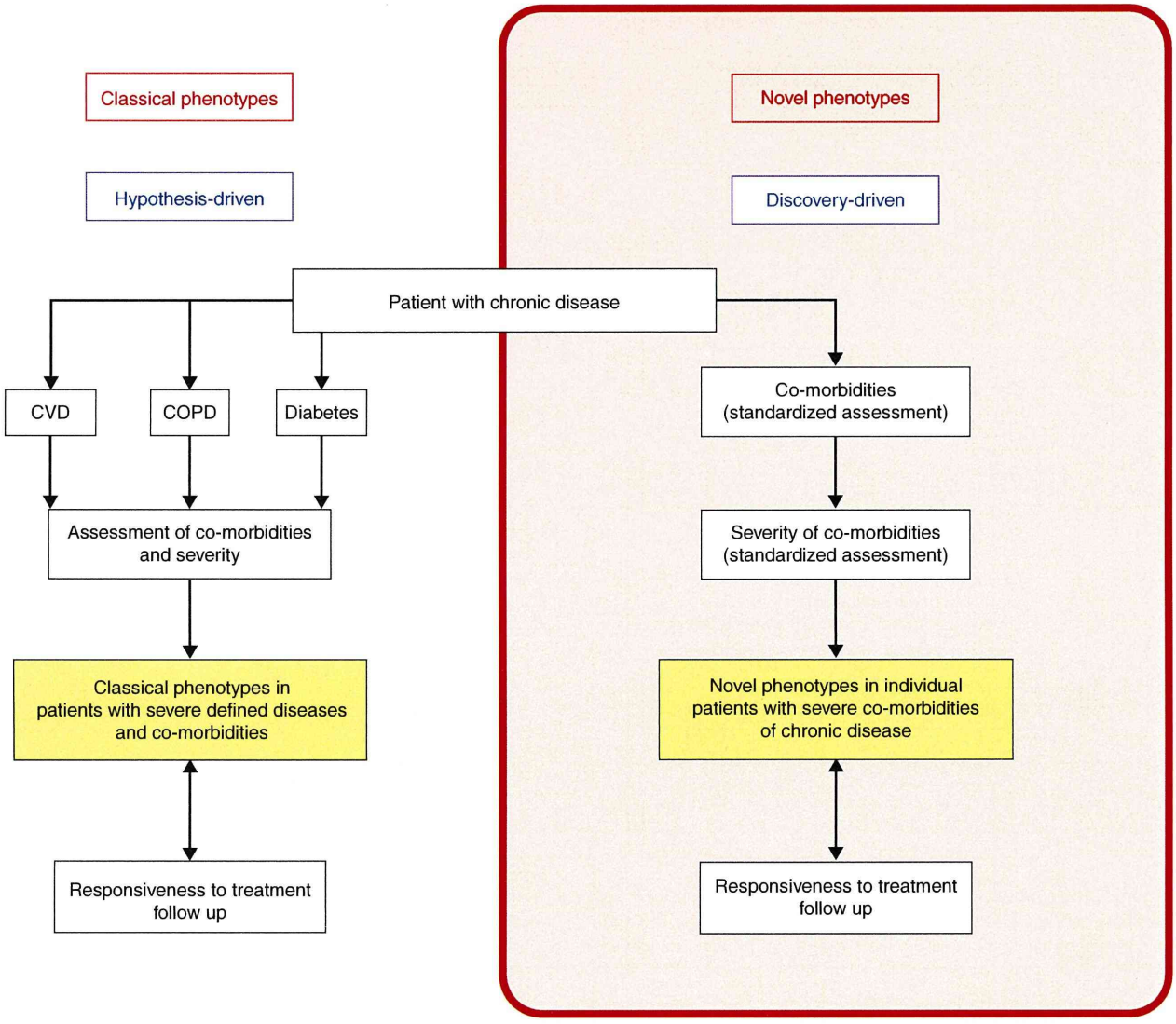


Figure 2. Classical phenotypes are based on *a priori* ontologies (cardiovascular disease, chronic obstructive pulmonary disease (COPD) and type 2 diabetes), and new phenotypes are based on statistical modeling of all the complex components of NCD onset, persistence and prognosis.

Information and communication technologies (ICT) are needed for the implementation of integrated care in a systems medicine approach to enable prospective follow-up of the patients. Home telemonitoring is promising [46] and should be explored further because continuous and precise monitoring makes each individual clinical history a valuable source of comprehensive information. More user-friendly and efficient ICT platforms are needed that include shared decision making, the process by which a healthcare choice is made jointly by the practitioner and the patient [47]. Ideally, an innovative patient management program would combine ICT, shared decision making and personalized education of the patient (and caregiver) about multidisciplinary approaches. The content, acceptance and effectiveness of such approaches should be tested to ensure that the autonomy, quality of life and capacity of patients are respected and enhanced, and that their values and preferences dominate decision making [48]. Practice-based inter-professional collaborations is also key to improving healthcare processes and outcomes [49]. Qualitative assessment will provide insight into how interventions affect collaboration and how improved collaboration contributes to changes in outcomes.

Thus, we propose that NCD management should move towards holistic multi-modal integrated care, and

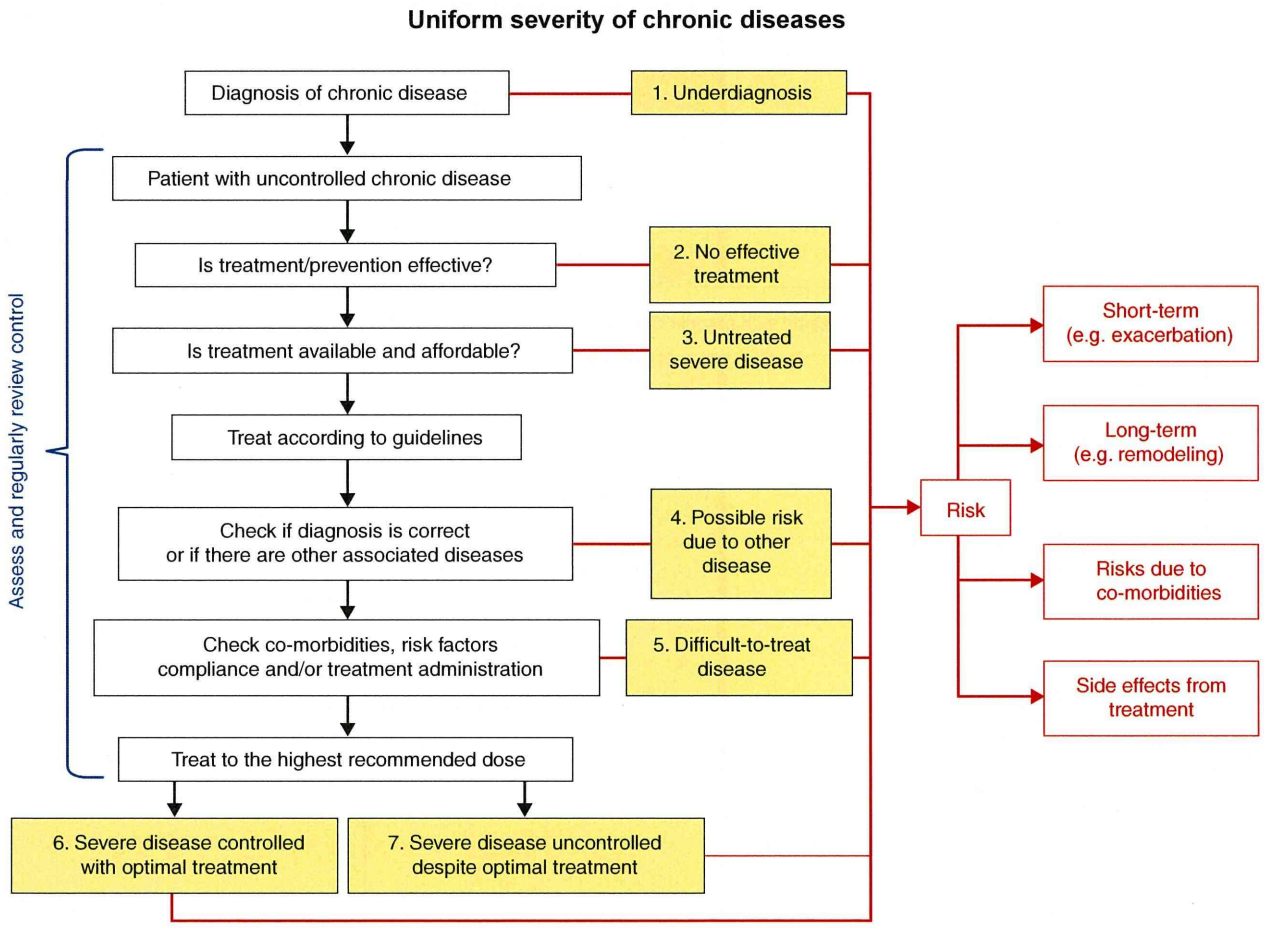


Figure 3. The concept of a uniform definition for NCD severity is based on control, responsiveness to treatment and risks (short, medium and long term). A single flow chart is proposed to define severity to improve phenotype characterization for all purposes (research, clinical and public health). It is based on diagnosis, therapeutic interventions and their availability/affordability, risk factors and co-morbidities.

multi-scale, multi-level systems approaches. To reduce their socio-economic and public health impacts, we propose that NCDs should be considered as the expression of a continuum or common group of diseases with intertwined gene-environment, socio-economic interactions and co-morbidities that lead to complex phenotypes specific for each individual. The 'systems medicine' concept, which takes a holistic view of health and disease, encapsulates this perspective. Systems medicine aims to tackle all components of the complexity of NCDs so as to understand these various phenotypes and hence enable prevention (Box 2), control through health promotion [50] and personalized medicine [51], and an efficient use of health service resources [52]. It does this through integrated care using multidisciplinary and teamwork approaches centered in primary and community care [53], including the essential ethical dimension.

Systems biology and medical informatics for P4 medicine of chronic NCDs

The main challenge regarding NCDs in the 21st century is to understand their complexity. Biology and medicine may be viewed as informational sciences requiring global systems methods using both hypothesis-driven and discovery-driven approaches. Systems medicine is the application of systems biology to medical research and practice [54,55]. Its objective is to integrate a variety of data at all relevant levels of cellular organization with clinical and patient-reported disease markers. It uses the power of computational and mathematical modeling to enable understanding of the mechanisms, prognosis, diagnosis and treatment of disease [56]. It involves a transition to predictive, preventive, personalized and participatory (P4) medicine, which is a shift from reactive to prospective medicine that extends far beyond what is usually covered by the term personalized medicine

Box 2: Glossary of terms

The classical definition of prevention [101] includes:

- **Primary prevention:** to avoid the development of disease.
- **Secondary prevention:** recognize a disease before it results in morbidity (or co-morbidity).
- **Tertiary prevention:** to reduce the negative impact of established disease by restoring function and reducing disease-related complications.

Expanding on the traditional model of prevention, Gordon [104] proposed a three-tiered preventative intervention classification system on the basis of the population for whom the measure is advisable based on a cost-benefit analysis:

- **Universal prevention** addresses the entire population (for example, national, local community, school, and district) and aims to prevent or delay risk factor exposure. All individuals, without screening, are provided with information and skills necessary to prevent the problem.
- **Selective prevention** focuses on groups whose risk of developing problems is above average. The subgroups may be distinguished by characteristics such as age, gender, family history, or economic status.
- **Indicated prevention** involves a screening process.

According to these definitions, **health promotion** [50] should be used for primary universal and selective prevention strategies, whereas **P4 medicine** (predictive, preventive, personalized and participatory) [51] should be used for primary, secondary and tertiary indicated prevention strategies.

Box 3: Key expected benefits of P4 medicine

To prevent the occurrence of NCDs by implementing effective action at societal and individual levels:

- To detect and diagnose disease at an early stage, when it can be controlled effectively.
- To stratify patients into groups, enabling the selection of optimal therapy.
- To reduce adverse drug reactions through the predictive or early assessment of individual drug responses and assessing genes leading to ineffective drug metabolism.
- To improve the selection of new biochemical targets for drug discovery.
- To reduce the time, cost, and failure rate of clinical trials for new therapies.
- To shift the emphasis in medicine from reaction to prevention and from disease to wellness.

[57,58]. It incorporates patient and population preferences for interventions and health states by implementing effective societal actions [57] with an important public health dimension [59]. It is likely to be the foundation of global health in the future (Box 3).

Thus, there is an urgent need for development of information management systems that can enable secure storage of heterogeneous data, including clinical data, and provide tools for the management, search and sharing of the data. Such information needs to accessible, shared between investigators, queried, and integrated in a controlled and secure manner with molecular profiles and images obtained from high-throughput facilities. For example, one prediction arising from considerations of the evolution of P4 medicine suggests that, in 10 years or so, each patient will be surrounded by a virtual cloud of billions of data points; we will need information technology to reduce this staggering data dimensionality to simple hypotheses about health and disease for each individual patient [57].

A systems biology approach that is unbiased by old classification systems can be used to find new biomarkers of co-morbidities, disease severity and progression. In this approach, phenotypes of NCDs are analyzed in an integrative manner using mathematical and statistical modeling, taking all diseases into account, and embedding co-morbidities, severity and follow-up of the patients through analyses in dynamic models (Figure 4). Unknown phenotypes are defined and further analyzed using iterative cycles of modeling and experimental testing. Novel biomarkers are identified combining datasets from genomics, epigenetics, proteomics, transcriptomics, metabolomics and metagenomics. These new complex biomarkers will need to be validated and replicated in independent controls or prospective patient cohorts [60]. Using methods used in non-medical complex model systems, it should be possible to monitor 'early warning signals', which predict the state of disease progression, and the occurrence of abrupt phase transitions (slowing down, increase in autocorrelation and variance) [61]. For example, in a mouse model of neurodegenerative disease, blood biomarkers have been shown to allow pre-symptomatic diagnosis and analysis of the stage of disease progression [62].

Modeling is a powerful tool for reducing the enormous complexity of comprehensive biological datasets to simple hypotheses. Modeling of the temporal behavior of disease read-outs at short [63] or long [64] intervals can identify sub-phenotypes of NCDs. Attempts to find novel biomarkers of disease development using a systems biology approach have been used to assess the mechanisms of severe asthma, allergy development [65] and cancer. One important role that biomarkers will have is to stratify a given disease into its different subtypes so that appropriate and distinct therapies can be selected for each subtype. Phenotypes can be modeled using statistical approaches, such as scale-free networks and Bayesian clustering models, that are based on the evaluation of NCDs as a whole, taking into account co-morbidities, severity and follow-up. This approach will

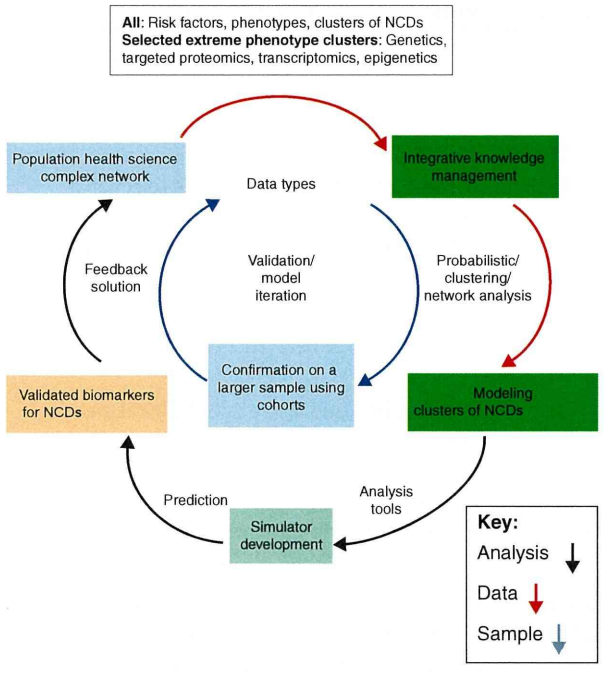


Figure 4. Iterative mathematical modeling to increase knowledge on NCDs. Various targeted or comprehensive data types are collected from samples of individuals with carefully defined phenotypes, processed using probabilistic and network analysis tools, and integrated into predictive models using a knowledge management and simulation platform, leading to the refinement of the classification of NCDs. Mechanistic hypotheses and complex biomarkers of NCDs generated through this process are then tested and validated iteratively using small then large cohorts of independent samples, providing potential diagnostic and therapeutic solutions for the general population.

make it possible to find intermediate phenotypes and patient-specific phenotypes. The challenge will be to develop efficient, automated and integrated workflows that predict the most suitable therapeutic strategy not only at the population level but, most importantly, at the individual patient level.

Bioinformatics, medical informatics and their interplay (sometimes termed biomedical informatics) will be key enablers in structuring, integrating and providing appropriate access to the enormous amount of relevant data and knowledge [66,67]. Medical informatics needs to provide ubiquitous and powerful electronic healthcare record technologies to securely aggregate and handle diverse, complex, and comprehensive data types [68]. Biomedical informatics must develop ways to use these content-rich electronic healthcare records to provide advanced decision support that considers all aspects of normal and disease biology, guided by clinically relevant insights and biomarker discovery research strategies [69,70]. Bioinformatics will need to constantly restructure and refine global data to distill the clinically useful elements and the derived models, so they can feed this information system in a real-time, automated fashion, constantly incorporating clinical expertise. P4 medicine is evolving so rapidly in its understanding of disease states that the individual patient's data must continually be re-examined so that new insights into the health and disease state of the individual can be gained. This general informatics framework, based on an advanced ICT infrastructure, will provide the basis for empowering P4 medicine.

Given the complexity of NCDs, bio-clinical scientific progress will depend critically on large-scale pooled analyses of high quality data from many biobanks [71] and bio-clinical studies (such as BioSHaRE-EU [72]). Biomedical informatics and knowledge management platforms have made significant advances towards enabling the development of technologies to organize molecular data at the level required for the complexity of NCD data [73,74]. Data analysis, integration and modeling require strict statistical procedures in order to avoid false discoveries [75]. They can be performed, for example, using the joint knowledge management platform of European Framework Program 7 (EU FP7) projects, including U-BIOPRED [76], MeDALL [65], AirPROM and SYNERGY-COPD, and using similar initiatives worldwide. Large-scale profiling to discover early markers of disease progression before the appearance of any symptoms has already been performed in a large prospective human cohort [77,78].

Complementary approaches using computational models that extend existing models derived from the Physiome project, including biomedical imaging, can be used together with statistical modeling of various types of clinical data to further define phenotypes and develop predictive models. These can be used within the framework of a fully integrated (preferably open source) knowledge management platform [79]. Such a platform for knowledge management, including annotation and ontologies, would then operate on top of the medical informatics infrastructure, setting the stage for a systems medicine approach to NCDs. In our collective experience these necessary aspects of medical informatics have a tendency to be overlooked in funding efforts targeting complex diseases.

Integrated care of chronic NCDs using P4 systems medicine

Integrated care, a core component of health and social care reforms, seeks to close the traditional gap between health and social care [80]. Population health sciences should integrate personalized medicine in public health interventions to prevent and manage NCDs in a cost-effective manner by involving all stakeholders, including patients [81]. The objectives of this proposed integration

are: (i) to investigate questions related to NCDs; (ii) to improve the quality of primary care; and (iii) to widely disseminate new information that will improve overall health at both a local and national level [82]. Chronic diseases can disconnect individuals from their usual milieu, with negative implications for physical, social and mental well-being. Moving beyond the disease-by-disease approach to tackle NCDs demands an improved understanding of NCD by patients, and a better understanding of their common causes. At the local level, strategies such as community oriented primary care can link and reinforce personal and public health efforts [83].

To understand, preserve and improve the health of human populations and individuals, an integrated research strategy should include all components of research on NCDs and be integrated for optimal patient management [84,85]. Careful evaluation is needed of: (i) the acceptance of multi-morbidity of NCDs by the patient, with particular attention to cultural and social barriers, gender and age; (ii) the engagement of patients in decisions regarding management [86], research and clinical trials [55,57]; and (iii) the improvement of quality of life that would result from the proposed management. Targeting NCDs and their comorbidities will directly affect healthy aging, which has been described as a 'keystone for a sustainable Europe' [87]. Screening, early diagnosis, prevention and treatment of hidden comorbidities in patients with diagnosed NCDs will reduce their morbidity and increase healthy life years.

The direct and indirect costs of uncontrolled NCDs are substantial for the patient, the family and society, especially in underserved populations [9]. P4 medicine should be put into the context of health economics to show that expensive strategies are cost-effective [55,57]. Chronic diseases place a considerable economic burden on the society and increase inequities. The social dimension of NCDs needs to be pursued in the economic and employment fields. The net social benefit of improving medical and social care related to NCDs should take co-benefits into account. Health costs for NCDs should be balanced with health benefits, wealth creation and economic development. The management of NCDs requires the coordination of stakeholders in the public and private sectors within a governance framework that includes networks of care. Therefore, research should be done to identify social determinants and to create public health systems that translate efficacy into effectiveness in the community [88]. Moreover, strengthening health equity across nations and socioeconomic groups is needed to meet the ambitions of the Commission on Social Determinants of Health, who have proposed closing the health gap between nations and groups in a generation [89].

Values are the basis of most actions in health and the economy, and these values are often not made explicit. Changing paradigms and approaches to NCDs may challenge fundamental societal values and professional habits [59,90]. The apparent contradiction between the development of a more tailored medical approach to NCDs and the public health dimensions of their prevention and care needs to be addressed using a value-based analysis. Thus, a thorough analysis of values underlying P4 medicine should be conducted in diverse contexts and should become part of the basis of decision-making. The respective weight of the multiple stakeholders involved in the priority setting must be made clear, with transparency and proportionality as key features. P4 medicine development should be a global aim and not a privilege of 'rich' countries. Using data obtained from all components of research, guidelines on NCDs applicable to primary care could be developed using up-to-date methodology [91,92]. Policies for implementation could then be proposed, to translate the concept of NCD into practice. They should distribute the burdens equitably, also considering gender and age.

Multidisciplinary training of all stakeholders, with particular emphasis on the participation of patient associations, is a further essential component. Many health and non-health professionals need to be educated in the general approach to the research and management of patients with NCDs. Innovative training programs using ICT will be essential in this implementation. Such education will also need to address questions of how to teach the subject and how people learn it, rather than merely regarding education as a process of transmission and transaction for everyone involved. This includes taking into account points of view, habits of mind, and all the information requested for the needs of the strategy. The educational program needs to forge educational systems to help participants think in a coherent way about NCDs. A module of the program should be developed around patient feedback to help them be engaged in all aspects of NCDs, including research.

Many patients with NCDs live in developing countries where medications and services are often unavailable or inaccessible. Effective medications, such as inhaled corticosteroids for asthma [93] or insulin for diabetes, should be made available for all patients [94]. In addition, there should be a global cost-effective application of P4 medicine across the world [95]. It is likely that genomic applications and ICT will become available to many developing countries at a relatively low cost in the next few years. In addition, new private-public strategic partnerships, such as the pre-competitive Innovative Medicines Initiative, a joint undertaking of the European Union and the European Federation of Pharmaceutical Industry Associations [96], and the Program on

Public-Private Partnerships of the United States National Institutes of Health Roadmap [97], are required to overcome the bottlenecks in the development of new treatment strategies [98]. WHO actively supports capacity building, especially in developing countries, fosters partnerships around the world, and works to narrow the gap in healthcare inequities through access to innovative approaches that take into account different health systems, economic and cultural factors. Despite the growing consensus for the need for health system strengthening, there is little agreement on strategies for its implementation [99]. Widely accepted guiding principles should be developed with a common language for strategy development and communication for the global community in general [100] and for NCDs in particular.

Conclusions

NCD management needs to move towards integrated care, global strategies and multi-modal systems approaches, which will reduce the burden and societal impact of NCDs. To this end, we propose that NCDs must be considered as the expression of a common group of diseases with different risk factors, socio-economic determinants and co-morbidities. This will enable the application of P4 medicine principles to NCDs, exploiting their commonalities, bringing improved global healthcare and the reduction of inequities around the world. The expected results targeted to better support for patients include: (i) better structuring of translational research and development for NCDs; (ii) greatly enhanced prevention and treatment capabilities; (iii) innovative healthcare systems with implementation of follow-up procedures directly in the homes of patients; (iv) slowing down of health expenditure increase; and (v) new interdisciplinary training curricula.

Abbreviations

AIRPROM, AIRway disease, PRedicting Outcomes through patient specific computational Modeling (FP7); BioSHare-EU, Biobank Standardization and Harmonization for Research Excellence in the European Union (FP7); ICT, information communication technology; MeDALL, Mechanisms of the Development of ALLergy (FP7); NAEPP-EPR3, National Asthma Education and Prevention Program, Expert Report 3; NCD, non-communicable disease; P4, predictive, preventive, personalized and participatory; U-BIOPRED, Unbiased BIOmarkers in PREDiction of respiratory disease outcomes (FP7); UN, United Nations; WHO, World Health Organization.

Competing interests

The authors declare that they have no competing interests in relation to the content of this article.

Acknowledgements

Part of the conceptual work presented has received support from the European commission FP7 projects AIRProm (Grant Agreement FP7 270194), BioSHaRE-EU (Grant Agreement FP7 261433), MeDALL (Grant Agreement FP7 264357), SYNERGY-COPD (Grant Agreement) and U-BIOPRED (Grant Agreement IMI 115010). JB, JMA, AC-T, FK, MLK, SP, CP and CA were supported by MeDALL; PJS, IMA and KFC were supported by U-BIOPRED; JR and AA were supported by SYNERGY-COPD; CB was supported by AIRProm; AC-T was supported by BioSHaRE-EU.

The positions, proposals and ideas expressed in this paper have been discussed by several authors (CA, ZC, LH, AB, JB, AC, SA, DC, DN) during the inaugural event of the European Institute for Systems Biology and Medicine of the Systemoscope International Consortium at the Biovision World Life Sciences Forum in Lyon on 28 March 2011.

Author details

[1]Department of Respiratory Diseases, Arnaud de Villeneuve Hospital, CHU Montpellier, INSERM CESP U1018, Villejuif, France. [2]Centre for Research in Environmental Epidemiology, Municipal Institute of Medical Research, Epidemiologia y Salud Publica, Universitat Pompeu Fabra, Doctor Aiguader, 88, E-08003 Barcelona, Spain. [3]Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands. [4]Cellular and Molecular Biology, Imperial College, South Kensington Campus, London SW7 2AZ, UK. [5]National Heart and Lung Institute, Imperial College, South Kensington Campus, London SW7 2AZ, UK. [6]Institut Clínic del Tòrax, Hospital Clínic, IDIBAPS, CIBERES, Universitat de Barcelona, Spain. [7]Department of Infection, Immunity and Inflammation, University of Leicester, Sciences Building, University Road, Leicester, LE1 9HN, UK. [8]Epidemiology, Public Health, Risks, Chronic Diseases and Handicap, INSERM U558, Toulouse, France. [9]IRCCS San Raffaele, Via della Pisana, 235, Rome, Italy. [10]Institut Pasteur, Bab Bhar, Avenue Jugurtha, Tunis, 71 843 755, Tunisia. [11]Division of Medical Genetics, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland. [12]Department of Diabetology, Montpellier, France. [13]Telethon Institute of Genomics and Medicine, Via Pietro Castellino, 111 80131 – Napoli, Italy. [14]Department of Pediatrics, University of Padova, Padova, Giustiniani, 3 – 35128, Italy. [15]Scientific Centre of Children's Health, Russian Academy of Medical Sciences, Lomonosovskiy prospect, 2/62, 117963,Moscow, Russia. [16]Department of Dermatology and Allergy, University of Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany. [17]Institut de Génomique Fonctionnelle, CNRS, UMR 5203, INSERM, U661, Université Montpellier 1 and 2, Montpellier, France. [18]Institute of Genomics and Integrative Biology, Near Jubilee Hall, Mall Road, Delhi-110 007, New Delhi, India. [19]Pulmonary Division, Albert Michallon University Hospital, Albert Bonniot Cancer Research Institute, La Tronche, Grenoble, France. [20]Endocrine Diseases, Lapeyronie Hospital, Montpellier, France. [21] Department of Physiology, Nîmes University Hospital, Place du Professeur Robert Debré. 30029 Nîmes Cedex 9, France [22]Department of Microbiology, Tumour and Cell Biology, Karolinska Institute, Nobels väg 16, KI Solna Campus, Box 280, SE-171 77 Stockholm, Sweden. [23]Department of Medical and Surgical Specialties, University of Modena and Regio Emilia, Modena, Italy. [24]Imperial College London, London, UK. [25]Institute for Systems Biology, Seattle, 401 Terry Avenue, North Seattle, WA 98109-5234, USA. [26]National Institute of Genetics, Mishima, Japan. [27]Auckland Bioengineering Institute, University of Auckland, Level 6, 70 Symonds Street Auckland, 1010. New Zealand. [28]Clinical Unit for Osteoarticular Diseases, and INSERM U844, Montpellier, France. [29]Centre for Research in Epidemiology and Population Health, INSERM U1018, Villejuif, France. [30]Singapore Immunology Network, 8A Biomedical Grove, Level 4 Immunos Building, 138648 Singapore . [31]Medical University of Lodz, Poland. [32]Department of Molecular Genetics, Weizmann Institute of Science, P.O. Box 26 Rehovot 76100, Israel. [33]Health Economy and Management, Paris-Dauphine University, Paris, France. [34]Biotechnology and Biotherapy, IRCM, Paris, France. [35]Department of Medicine, Groote Schuur Hospital and University of Cape Town, South Africa. [36]Department of Physiology, Montpellier University, and INSERM U1046, France. [37]The Estonian Genome Center of University of Tartu, Tartu, Estonia. [38]Epsylon, Montpellier, France. [39]Department of Physiology, University of Oxford, Le Gros Clark Building, South Parks Road, Oxford OX1 3QX, UK. [40]Department of Molecular Biology and Genetics, Bilkent University, Faculty of Science, B Building, 06800 Ankara, Turkey. [41]European Patient's Forum (EPF) and European Federation of Allergy and Airways Diseases Patients Associations (EFA), Brussels, Belgium. [42]Department of Medicine, University of Kiel, Germany. [43]Department of Family Medicine, University of Wisconsin, 1100 Delaplaine Ct. Madison, WI 53715-1896, USA. [44]Department of Public Health, AL. JEROZOLIMSKIE 87, 02-001 Warsaw, Poland. [45]Departments of Clinical Epidemiology and Biostatistics and of Medicine, McMaster University, 1280 Main Street West, Rm. 2C12, L8S 4K1 Hamilton, ON, Canada. [46]Institute of Pharmacology, University of Bern, Friedbühlstrasse 49, CH-3010 Bern, Switzerland. [47]Cancer Biology and Epigenomics Program, Children's Memorial Research Center and Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, USA. [48]Research Centre for Molecular Medicine, Lazarettgasse 14, AKH BT 25.3, A-1090, Vienna, Austria. [49]Department of Medicine, Karolinska Institute, Solna,