Human Molecular Genetics, 2008, Vol. 17, No. 21 ***3411–3425***
doi:10.1093/hmg/ddn235
Advance Access published on August 8, 2008

Human neural crest cells display molecular and phenotypic hallmarks of stem cells

Sophie Thomas[1], Marie Thomas[1], Patrick Wincker[2], Candice Babarit[1], Puting Xu[3], Marcy C. Speer[3,†], Arnold Munnich[1,4], Stanislas Lyonnet[1,4], Michel Vekemans[1,4] and Heather C. Etchevers[1,5,*]

[1]INSERM, U781, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75015 Paris, France, [2]GENOSCOPE (CEA) and UMR 8030 CNRS-GENOSCOPE-Université d'Evry, 2 rue Gaston Crémieux, 91057 Evry, France, [3]Center for Human Genetics, Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA, [4]Faculté de Médecine, Université Paris Descartes, 15 rue de l'Ecole de Médecine, 75005 Paris, France and [5]INSERM, U563, Centre de Physiopathologie de Toulouse-Purpan, 31300 Toulouse, France

Received June 23, 2008; Revised and Accepted August 6, 2008

The fields of both developmental and stem cell biology explore how functionally distinct cell types arise from a self-renewing founder population. Multipotent, proliferative human neural crest cells (hNCC) develop toward the end of the first month of pregnancy. It is assumed that most differentiate after migrating throughout the organism, although in animal models neural crest stem cells reportedly persist in postnatal tissues. Molecular pathways leading over time from an invasive mesenchyme to differentiated progeny such as the dorsal root ganglion, the maxillary bone or the adrenal medulla are altered in many congenital diseases. To identify additional components of such pathways, we derived and maintained self-renewing hNCC lines from pharyngulas. We show that, unlike their animal counterparts, hNCC are able to self-renew *ex vivo* under feeder-free conditions. While cross species comparisons showed extensive overlap between human, mouse and avian NCC transcriptomes, some molecular cascades are only active in the human cells, correlating with phenotypic differences. Furthermore, we found that the global hNCC molecular profile is highly similar to that of pluripotent embryonic stem cells when compared with other stem cell populations or hNCC derivatives. The pluripotency markers *NANOG*, *POU5F1* and *SOX2* are also expressed by hNCC, and a small subset of transcripts can unambiguously identify hNCC among other cell types. The hNCC molecular profile is thus both unique and globally characteristic of uncommitted stem cells.

INTRODUCTION

Widespread proliferation and progressive fate restriction over time characterize embryonic development. During and after neural tube closure in vertebrate embryos, neural crest cells (NCC) detach from the ectoderm at the boundary between neural and non-neural epithelia, multiply and infiltrate the mesoderm. The site-appropriate differentiation of NCC is the result of a combination of extrinsic factors from the embryonic niche (1,2) and cell-intrinsic properties that modify responsiveness to these influences (3). NCC normally yields neurons and glial cells of the entire peripheral nervous system (PNS), pigment cells and endocrine cells (4). In the head, they also give rise to cephalic tendons, cartilage, bone, dermis, meninges, vascular smooth muscle and adipocytes (5). The original progenitors disappear along with their birthplace as the neural tube closes and matures. However, such locations as the enteric ganglia (6), the dorsal root ganglia (7), the hair follicle (8–10), the tooth (11) and even the bone marrow (12,13) appear to be later niches for the maintenance of persistent, oligopotent avian and rodent neural crest-derived stem cells.

Because of its wide range of derivatives and long-term plasticity, the development of the neural crest has been the topic of

*To whom correspondence should be addressed. Tel: +33 562744587; Fax: +33 562744558; Email: heather.etchevers@inserm.fr
†Died August 4, 2007.

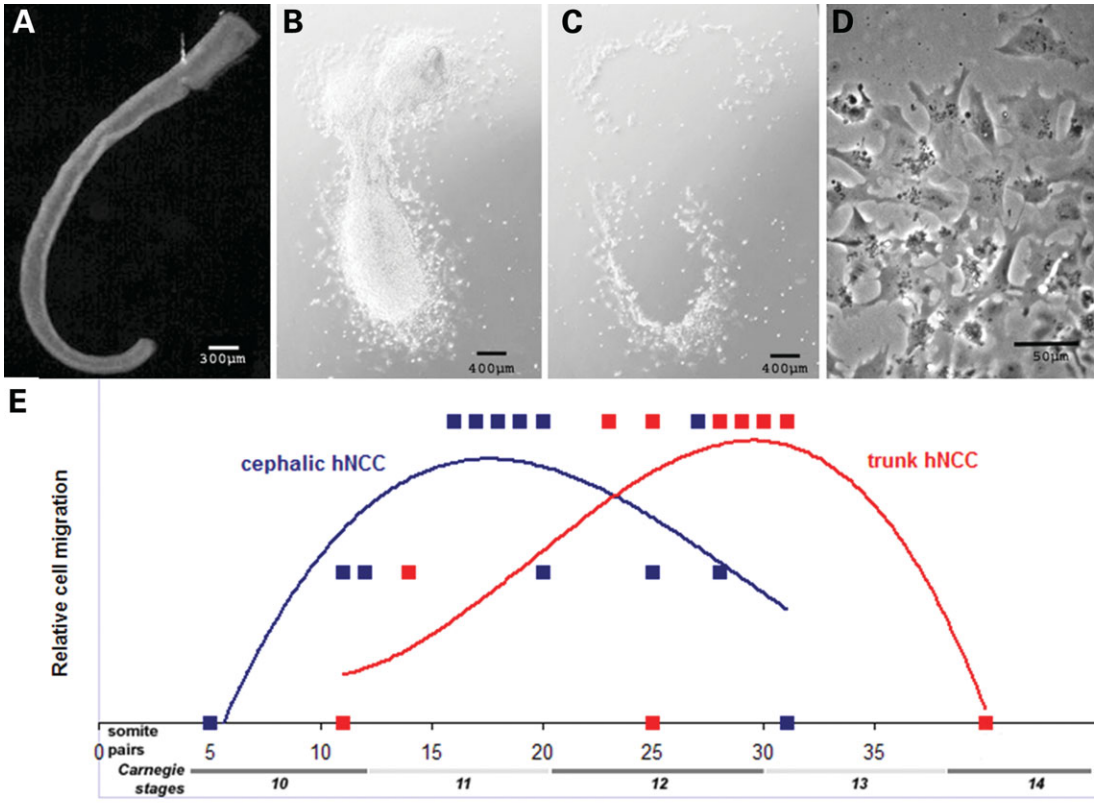

Figure 1. Primary neural crest cells can be isolated from human embryos. (**A**) Human neural tube (NT) from an embryo at Carnegie stage (C)13. (**B**) After 16 h, most hNCC have migrated away from the dorsal NT. (**C**) The intact NT is detached from the culture dish. (**D**) An enriched hNCC population remains, phenotypically similar to murine or avian NCC. (**E**) Empirical evaluation of hNCC migration (none to few, some and many) from 31 explanted neural tubes. Third-order polynomial regressions reflect the rostral-to-caudal temporal maturation gradient of the human NT and maturation of NCC. Peaks occur during C11 at cephalic levels, and during late C12 at rostral trunk levels (segments extending from somites 5 through the last-formed somite pair).

intense study in many vertebrate species. Almost nothing is known about the endogenous characteristics of human (h)NCC, as they mix with other cell types almost immediately. Genetic errors influencing hNCC development seem to be the basis of such common birth defects as congenital heart defects, Hirschsprung disease (HSCR), labiopalatine clefting or cancers such as neuroblastoma and pheochromocytoma, collectively known as neurocristopathies (reviewed in 14). We set out to identify molecular networks that were activated in an early hNCC population before they dispersed to their final sites of differentiation.

To address this issue, we first determined the precise time window during which hNCC separate from the developing neural tube. We then derived primary cell lines that self-renew without a feeder layer and can be propagated and frozen for many cycles. This had not been accomplished to date with NCC from any other species. Rather than focus on known pathways, we examined their entire transcriptome using SAGE to determine an intrinsic molecular profile. A subset of transcripts was validated using RT–PCR, immunohistochemistry and *in situ* hybridization to check representativity. When compared to published data from murine or avian counterparts, hNCC activate novel signaling pathways on top of many evolutionarily conserved modules. Furthermore, the hNCC transcriptional profile was highly evocative of the molecular signature of human embryonic stem (hES) cells, including but not restricted to the expression of the transcription factors *SOX2*, *NANOG* and *POU5F1*. These data indicated that after separation from the neuroepithelium, the plastic hNCC population remains poised to respond to lineage-inductive cues, using much the same transcriptional machinery as hES cells to delay differentiation.

RESULTS

Human neural crest cell isolation

In order to study hNCC before they migrated among unrelated cell types, we first determined the stages during which they emigrate from the dorsal neuroepithelium. To this end, 31 human neural tube fragments from embryos with normal trophoblast karyotype and ranging between Carnegie stages (C) 10 and C14 were isolated and plated on collagen I-coated culture dishes for 16 h (Fig. 1A) in conditions similar to those used in animal model systems (15,16). Cephalic neural tubes (from prospective diencephalon through the rhombencephalon) yielded a halo of hNCC spread onto the dish surface between late C10 to late C12 [23–27 days post-fertilization (dpf); Fig. 1B and C]. A similar halo formed around lengths of trunk-level neural tube caudal to somite 5 between the stages C11 and C13 (24–29 dpf), but not around neural tube fragments taken from outside these spatiotemporal boundaries (Fig. 1E). The neural

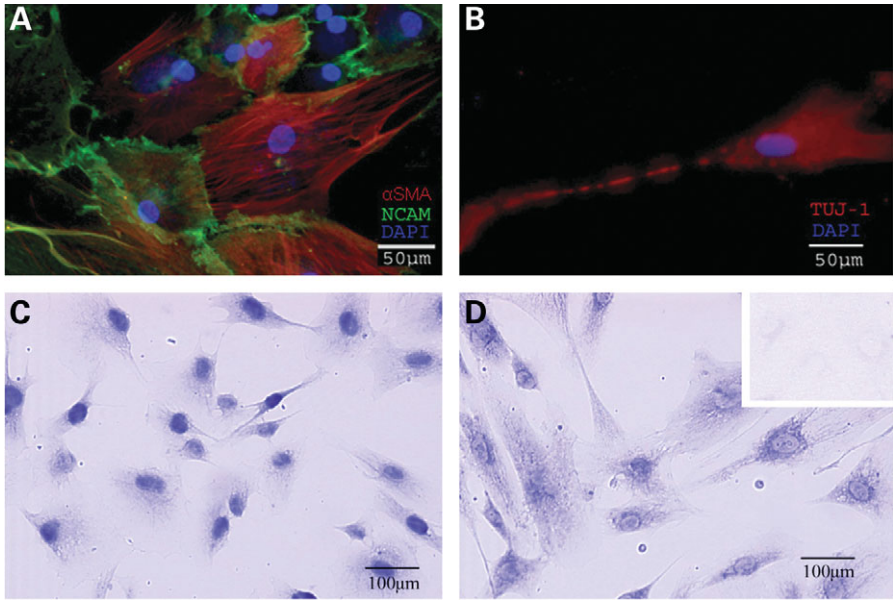

Figure 2. Protein markers of hNCC indicate a heterogeneously uncommitted phenotype. *In vitro*, hNCC synthesized (**A**) α-smooth muscle actin and/or NCAM, separately and sometimes in the same cells; (**B**) neuron-specific beta III tubulin; (**C**) nuclear SOX9; (**D**) nuclear and cytoplasmic SOX2; and unpolymerized GFAP (not shown). Pigment or immunoreactivity to calcitonin or tyrosine hydroxylase was never observed in these culture conditions. Inset: negative control with non-specific primary antibody.

tube was then detached and removed (Fig. 1C), leaving a highly enriched NCC population with a characteristic, fibroblast-like morphology (Fig. 1D). Four cephalic and eight trunk-level cell lines were derived and propagated. Cells remained euploid and morphologically unchanged after more than 20 passages and multiple freeze/thaw cycles. The population doubling time of hNCC cell lines is 40–48 h and does not change substantively after 9 months *in vitro*.

Using immunohistochemistry, we observed that some hNCC expressed proteins typical of partially committed progenitors in avian and murine NCC cultures, such as alpha smooth muscle actin, neural cell adhesion molecule (NCAM) (Fig. 2A) and glial fibrillary acidic protein (not shown). We never observed the hNCC derivative-specific markers tyrosine hydroxylase or calcitonin; however, occasional neurons, revealed by neuron-specific class III beta-tubulin (TuJ-1) staining, were present in cultures (Fig. 2B). The SOX9 protein, which is crucial for differentiation of the full range of NCC derivatives (17), was localized mostly in the nucleus (Fig. 2C). In contrast, the SOX2 protein, a marker of both stem cells and uncommitted neuroepithelium (18), was both nuclear and cytoplasmic (Fig. 2D).

These observations suggest that individual hNCC within a given cell line are poised at various stages of lineage commitment, as observed in animal models immediately after delamination from the neuroepithelium (19,20).

Serial analysis of gene expression and validation

For complete transcriptional characterization of hNCC, we used a modification of the Serial Analysis of Gene Expression (SAGE) technique (21) to generate 'long' 21 basepair tags, representing gene transcripts among the total RNA derived from trunk-level hNCC. These cells have been isolated from a C13 female embryo with a normal trophoblast karyotype and passaged seven times at the time of the analysis. Sequencing of 3546 clones yielded 50 500 tags after exclusion of duplicated ditags (GEO accession GSM207304). Of these, 22 797 were unique tags, representing 8831 transcripts with distinct UniGene identifiers. Of the unique tags, 39% could not be reliably assigned to known mRNAs.

To validate the SAGE data, we examined the expression of 55 genes identified as present in the hNCC library using semi-quantitative RT–PCR (Fig. 3; Supplementary Material, Fig. S1). Among the transcripts tested, 50 were represented by tags with very low (≤ 3) to low ($3 < n < 10$) abundance, or less than 0.02% of total transcripts. Nearly all (96.4%) were confirmed, including all 20 genes with three or fewer tags (Supplementary Material, Table S1).

Eighteen classical signature RNAs of amniote NCC (22,23) were expressed in five different trunk-level hNCC lines, including the one used for SAGE analysis (N5 in Fig. 3). These included transcription factors such as *FOXD3*, *MSX1*, *SNAI2*, *SOX9*, *SOX10* and *TWIST*, as well as signaling molecules or membrane-bound receptors such as *RET*, *GJA1*, *EDN1*, *EDNRB*, *NESTIN*, *NOTCH1, P75*, *PDGFA*, *PDGFB* and *PDGFRB* (Fig. 3A).

All transcripts tested were present in the five hNCC lines as well as in the stage C12–C13 trunk-level neural tubes from which the NCC emigrate, with the exception of *PAX3* and *PAX6*, absent from line N3. The embryonic C13 liver bud also expressed most of these genes. We examined a subset in cDNA from adult human liver, and again, many were detected (Fig. 3A). This expression was not artefactual, as *PAX6* was not expressed by the liver bud, or *MSX1* by the adult liver. Transcripts, such as *PAX3*, *FOXD3*, *NES* and *TWIST*, appeared more abundant in hNCC than in the embryonic liver, unlike *GJA1* or *SMAD2*. We confirmed that the sensitivity of the semi-

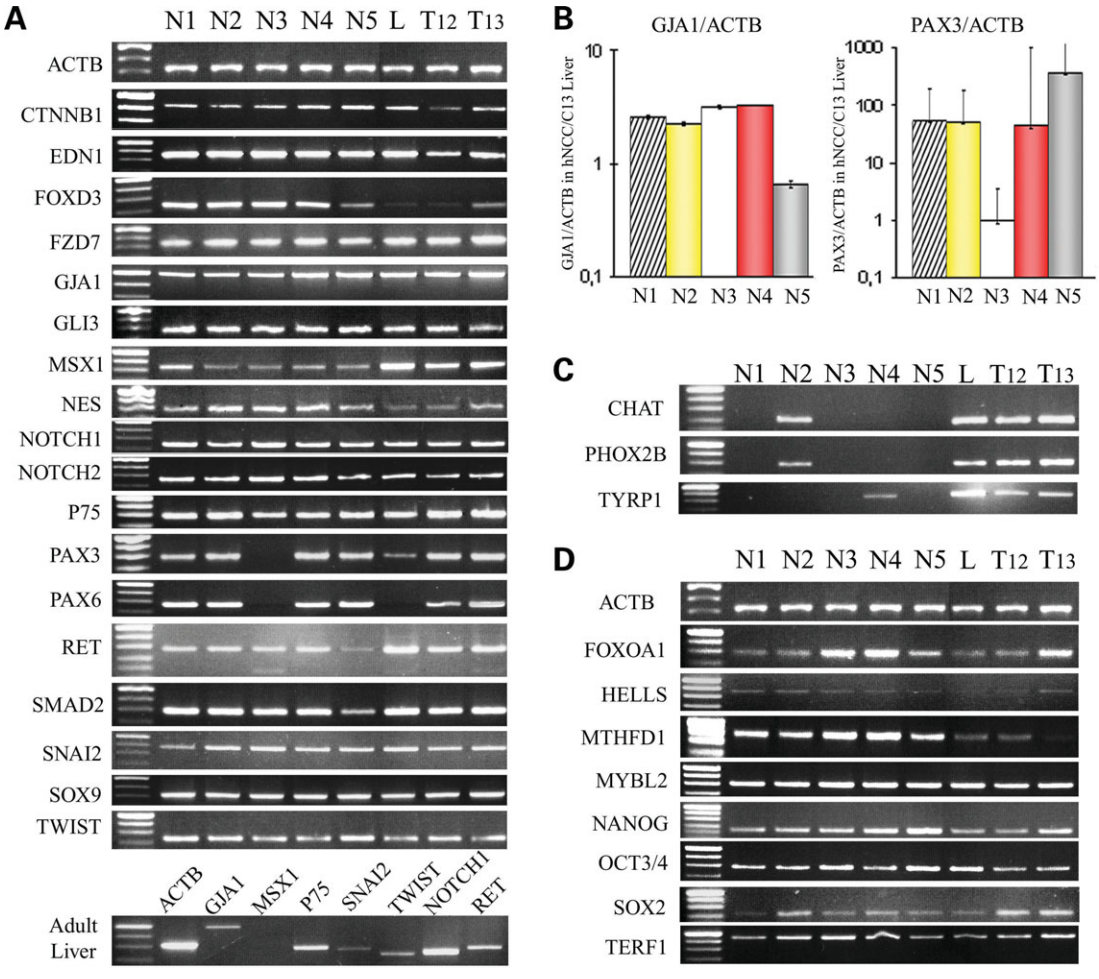

Figure 3. Expression validation of the hNCC SAGE bank. (**A**) Presence in five distinct hNCC lines (N1–N5) of typical animal NCC gene transcripts, as shown by RT–PCR. Most of these genes are also expressed in the C12 and C13 human neural tube (T12 and T13) and in the C13 liver bud (L), as well as in the adult human liver, below. (**B**) *GJA1* (*CX43*) and *PAX3* are expressed more in hNCC than embryonic liver. *ACTB* expression was used to normalize the data before calculating the expression ratio of each gene in the five hNCC lines compared to the C13 liver bud. (**C**) Choline acetyltransferase (*CHAT*) and paired-like homeobox 2b (*PHOX2B*) are expressed by sympathetic neurons and tyrosinase-related protein 1 (*TYRP1*) by melanocytes, both hNCC derivatives: none of the five hNCC lines express all three markers while T12/T13/L samples do. (**D**) Expression in the five distinct hNCC lines of genes reported in the literature as highly characteristic of hES cells.

quantitative approach masked discrete differences in expression levels by quantitative RT–PCR. *PAX3* transcripts were at least 50-fold, and *GJA1* at least 3-fold, more abundant in hNCC compared to the embryonic liver bud, when normalized to *ACTB* abundance (Fig. 3B). In addition, three genes unrepresented in the hNCC SAGE bank and characteristic of differentiated hNCC progeny (Fig. 3C, *CHAT* and *PHOX2B*, sympathetic neurons; *TYRP1*, melanocytes) were not amplified by semi-quantitative RT–PCR in the original hNCC line used (N5). However, the sympathetic markers were detectable in line N2 and *TYRP1* in line N4.

We finally examined the spatial expression of a selection of genes identified by SAGE using *in situ* hybridization on human embryo sections at C13 (Fig. 4). *SOX11* and *MAZ* code for transcription factors and *GJA1* for a critical gap junction protein; other genes we studied (not shown) include the transcription factors *SOX10* (24), *ZEB2* (25) and *CHD7* (26) and *HEYL*; the receptors encoded by *NOTCH2* and *FGFR2*; and the cytoskeleton-associated *CTNNB1* and *MID1* (27). All were expressed in both the neuroepithelium and NCC, with the exception of *SOX10*, which only postmigratory hNCC appeared to express at C13. In addition, *SOX11* (Fig. 4B), *GJA1* (Fig. 4C, F and I) and *MAZ* (Fig. 4D) were transcribed by mesodermal derivatives: *GJA1* by the pronephros, limb bud mesenchyme and, like *SOX11* and *MAZ*, the dermamyotome. Endodermal epithelia expressed *FGFR2*, similar to later stages (28). All genes tested were, therefore, expressed by other tissues, notably by the source neuroepithelium, in addition to hNCC at C13. As in animals, *SOX10* (24) and *FOXD3* (Fig. 3) appeared to be more expressed by early postmigratory hNCC than the neural tube.

Overall, these results show that this hNCC SAGE bank accurately reflects the gene activity of line N5. Its transcriptome is generally representative of hNCC isolated in this manner, although quantitative expression level differences indicate that hNCC lines may be individually heterogeneous, as for other stem cell lines. Moreover, the *in situ* data confirm that gene transcripts that are present in the N5

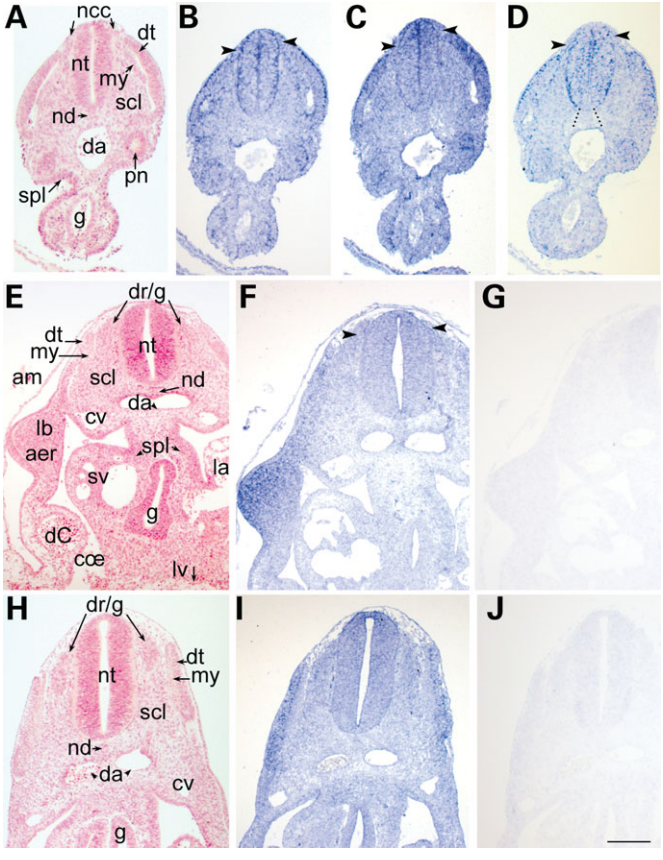

Figure 4. *In situ* expression of *SOX11*, *MAZ* and *GJA1* in the human embryo. (**A**) Hematoxylin-eosin (HE) stain of caudal trunk-level transverse section at stage Carnegie (C)13. (**B**) Adjacent section. Presumptive dermatome, neural tube excluding floorplate, pronephros and splanchnopleural mesoderm and neural crest cells (arrowheads) express *SOX11*. (**C**) The same tissues express *GJA1* in a simultaneously hybridized adjacent section. Expression is higher in the dermatome, dorsal neural tube and migratory neural crest cells (arrowheads). (**D**) *MAZ* is expressed in an adjacent section in most of the neural tube aside from the floorplate (indicated as extension of dotted lines), and lightly in neural crest (arrowheads), dermamyotome and pronephros. (**E**) Forelimb-level HE transverse section in separate C13 embryo. (**F**) *GJA1* antisense probe-hybridized adjacent section. Limb bud, dorsal neural tube and dorsal roots (arrowheads) hybridize more strongly than other tissues, nearly all of which have some basal *GJA1* expression. (**G** and **J**) *GJA1* sense probe-hybridized adjacent section to previous frame, demonstrating specificity of antisense hybridizations. (**H**) Rostral trunk-level HE section of same embryo as in (A–D). (**I**) *GJA1* is highly expressed in the dorsal neural tube, the dorsal roots and ganglia, the dermatome and the splanchnopleural mesoderm, with basal expression in all other tissues of the section as compared with (J). Abbreviations: aer, apical ectodermal ridge; am, amnion; cœ, cœlom; cv, cardinal vein; da, dorsal aorta; dC, duct of Cuvier; dr/g, dorsal root more or less in plane of ganglion; dt, dermatome; g, gut; la, left atrium; lb, limb bud; lv, liver bud; my, myotome; ncc, neural crest cells; nd, notochord; nt, neural tube; pn, pronephros; scl, sclerotome; spl, splanchnopleura; sv, sinus venosus. Bar = 250 μm.

hNCC SAGE bank after cell enrichment *in vitro* are representative of hNCC gene expression *in vivo*.

Interspecies NCC transcriptome comparisons

Most genes that are commonly used to characterize NCC in animal studies are expressed by hNCC as well. Of the 58 transcripts identified in an early survey of avian NCC cDNAs (23), 95% were also observed in the hNCC SAGE bank; some of these have many human paralogues (Supplementary Material, Table S2). A more comprehensive effort using SAGE on mouse NCC after 2 (mNCSC) and 7 (mNCP) days of culture (22) identified approximately 6000 murine gene transcripts. Of these, 67.2% have orthologues in hNCC (Fig. 5A). However, more than 4000 additional Unigene clusters were present only in the human cells. This correlates with such phenotypic differences observed *in vitro* as the adherent human cell independence from a feeder layer and unlimited ability (to date) for propagation without large-scale arrest and differentiation.

The IDEG6 web tool (29) was applied and yielded a list of more than 6000 genes that were differentially regulated in hNCC with statistical significance compared to the combined gene list of mNCSC/NCP. Ontological analysis of the differential group of genes indicated that the most enriched functional categories in hNCC were cell signaling, cell death, gene expression, cellular growth and proliferation and cell-cycle regulation (Fig. 5B). Furthermore, genes annotated as functioning within the insulin, Shh, Wnt and other growth factor signaling pathways were enriched in the human cells (Fig. 5C and D).

Cluster and pathway analyses of whole transcriptomes show similarity between hNCC and embryonic stem cells

In order to assess the global functional significance of the hNCC gene list relative to a variety of other human cell and tissue types, we undertook average linkage hierarchical clustering of the gene lists, using SAGE transcriptome data from 14 other non-transformed samples. This analysis grouped hNCC at a distance from tissues derived from other germ layers, such as the kidney or skeletal muscle, but also from tissues of embryologically closer lineages, such as adult Schwann cells (30) or most of the central nervous system. The closest clustering was observed with hES cell lines hES3 and hES4 (31), in contrast to human mesenchymal stem cells (32) (Fig. 6A). One interpretation of these observations is that transcriptomes cluster cell or tissue types by progenitor commitment rather than embryological origin.

In line with this interpretation, hNCC express many genes considered to play essential roles in the maintenance of multi-/pluripotency. *NANOG*, *POU5F1* (*OCT3/4*) and *SOX2* (33–35), considered emblematic of embryonic stem cells, are all synthesized by hNCC (Fig. 3D). In humans, we observe widespread transcription of *NANOG* and *POU5F1* at pharyngula stages (Fig. 6B–D). Expression remains high in central and PNSs at C15 (Fig. 6E–G) and C17–18 (Supplementary Material, Fig. S2), but decreases dramatically in most other somatic tissues. Many other established stem cell markers such as *CD24*, *FOXH1* and *LIN28* are transcribed by hNCC (Fig. 3D), as are *MYBL2*, *HELLS*, *EPHA1*, *GPR23* and *PHC1* (36).

SOX2, *NANOG* and *POU5F1* co-occupy and appear to positively regulate 180 promoter targets in hES cell line H9 (37). Of these targets, 110 are expressed by hNCC, when compared with the 98 transcribed by hES3 and/or hES4 (Supplementary Material, Table S3). The comparison of quantitative transcriptional levels of common co-occupied targets between hNCC

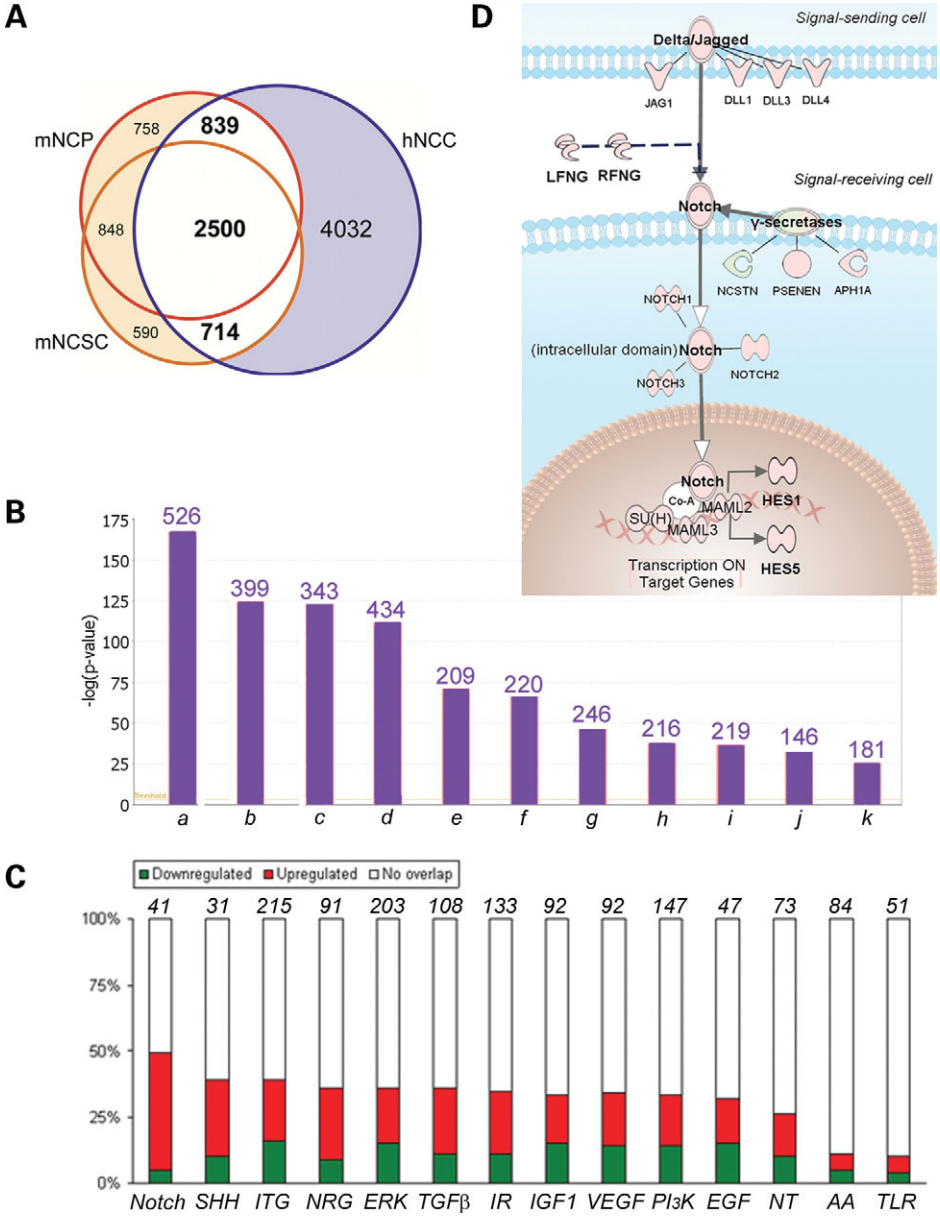

Figure 5. Human and mouse neural crest transcriptomes have much in common but are also species-specific. (**A**) Venn diagram with common and specific genes to hNCC, mouse (m)NCSC (early cultures) and mNCP (later cultures) as described by Hu *et al*. (22). (**B** and **C**) Functional annotation of those hNCC genes differentially expressed ($P < 0.001$; Fisher's exact *t*-test with Benjamini–Hochberg correction for multiple testing) with respect to the combined set of m(NCSC+NCP). (**B**) Statistically over-represented functional groups in hNCC with number of molecules assigned to a given group over each bar. (**C**) Under- and over-expression of genes assigned to individual pathways in hNCC relative to the mouse represented in green and red, respectively. White represents those members of a category absent from one or the other dataset. (**D**) Schematic view of individual components of part of the Notch pathway from (C) with the same color convention, as expressed in hNCC. In contrast to mNCSC/mNCP, hNCC express many Notch ligands, receptors, co-activators, effectors and transcriptional targets. Abbreviations: a, cell signaling; b, cell death; c, gene expression; d, growth and proliferation; e, cell cycle; f; cytokinesis; g, nervous system development and function; h, cell morphology; i, cell–cell interaction; j, embryonic development; k, hematological system development and function; AA, eicosanoids; EGF, epidermal growth factor; ERK, Microtubule-associated protein kinases; hNCC, human neural crest cells; IGF1, insulin-related growth factor 1; IR, insulin receptor; ITG, integrins; mNCP, mouse neural crest progenitors; mNCSC, mouse neural crest stem cells; NRG, neuregulins; NT, neurotrophins; PI3K, Phosphoinositide 3-kinases; SHH, sonic hedgehog; TGFb, transforming growth factor beta family; TLR, Toll-like receptors; VEGF, vascular endothelial growth factors.

and hES3/4 shows that levels of these genes are globally similar among these three cell lines but differ from bone marrow- or umbilical cord-derived mesenchymal stem cells, adult liver or lung (Fig. 6H). Finally, among 280 genes considered to be stem cell-specific after comparison of seven genetically independent hES cell lines (38), 120 are co-expressed by hES3/4 and hNCC (Supplementary Material, Table S4).

Identical broad-category molecular networks were statistically over-represented ($P < 0.001$) in hES and hNCC SAGE libraries (Supplementary Material, Fig. S3), although this

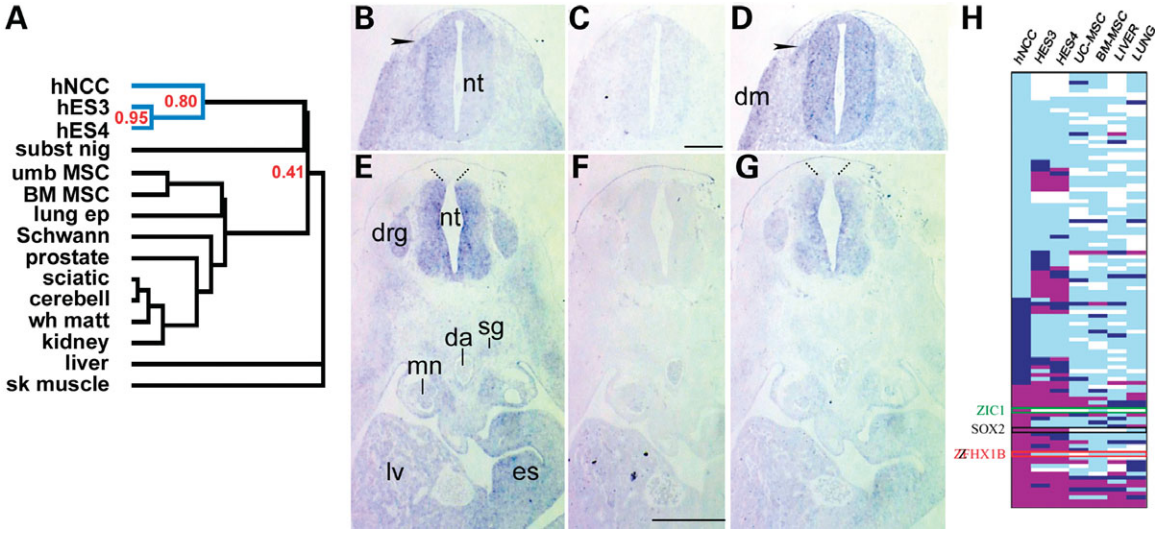

Figure 6. Pluripotent stem cell markers are expressed by uncommitted hNCC. (**A**) Hierarchical cluster dendrogram of hNCC-expressed transcript list compared to 14 publicly available, normal human tissue SAGE banks (uncentered correlation, average linkage). The global transcriptome of hNCC is most similar to two hES cell lines (hES3 and hES4) relative to the transcriptomes of the substantia nigra (subst nig); mesenchymal stem cells from the umbilical cord (umb MSC) or bone marrow (BM-MSC), highly similar to each other by this analysis; pulmonary epithelium (lung ep), Schwann cells *in vitro*, prostate, sciatic nerve, cerebellum, brain white matter (wh matt), kidney, liver or skeletal (sk) muscle. Correlation coefficients are indicated in red. (**B**) Expression of *NANOG* mRNA in transverse section of human Carnegie stage 13 (C13) embryo (cf. Fig. 4H) is discrete but present in neural tube, neural crest cells in the dorsal root (arrowhead) and dermatome, when compared with a sense probe hybridized adjacent section (**C**). (**D**) Expression of *POU5F1* in simultaneously hybridized adjacent section is more visible in equivalent structures, and seems to have a widespread basal expression in all tissues. (**E**) *NANOG* expression at C15 (cf. Fig. 9F and G). After a few days' growth, expression is more distinct in the proliferating neuroepithelium, dorsal root ganglia, the motor horns, stomach, sympathetic ganglia, liver bud and migrating myotome cells. (**F**) Adjacent section hybridized with sense probe. (**G**) *POU5F1* is expressed in a similar pattern. Both *NANOG* and *POU5F1* transcripts appear to be excluded from the roofplate of the neural tube at this stage (extension of dotted lines). (**H**) Heat map of target genes whose promoters can be co-occupied by *SOX2*, *NANOG* and *POU5F1* in hNCC, hES3, hES4, BM-MSC, UC-MSC, adult liver and lung cells, respectively. White stands for a tags-per-million (tpm) value equal to 0, light blue for $1 \leq$ tpm ≤ 49, dark blue for $50 \leq$ tpm ≤ 100 and magenta for tpm value >100. Global modulation is similar in hNCC and hES3/4 compared to the more distantly clustered cell types; for example, *SOX2* itself, although other genes are expressed differentially, such as *ZIC1 or ZFHX1B*. Abbreviations: da, dorsal aorta; dm, dermamyotome; drg, dorsal root ganglion; es, stomach; lv, liver; mn, mesonephros; nt, neural tube; sg, sympathetic ganglion.

observation mostly reflects the gross resolution of such ontologies. More detailed analysis showed that human NCC and ES cells synthesized 840 mRNAs in common that are annotated as involved in 'transcriptional regulation' (Supplementary Material, Table S5). Of these, 114 transcription factors (Supplementary Material, Table S6) included targets of characterized growth factor (*SMAD1/3/4/5*, *CREBBP*, *STAT1/3*, *SRF*, *JUN*, *CUTL1*, *ELK1* and *FOS*) and steroid hormone (*RARA*, *RXRA*, *RXRB*, *RARG*, *NR1D2* and *VDR*) signaling pathways. When transcriptional regulation genes were also expressed by the multipotent BM-MSC or the liver, there was wide fluctuation in expression levels between cell types, which was not the case for the hNCC–hES comparison.

The second most over-represented classification of gene products common to hNCC and hES was that of 'cellular proliferation and growth'. Within the category, hNCC, hES3 and hES4 expressed 1088 genes in common (out of 1350, 1712 and 1601, respectively, in this category) that regulate cell cycle progression and encode growth factors, hormone and cytokine receptors and their effectors (Supplementary Material, Table S7). In all stem cell lines, more members of the hedgehog, fibroblast growth factor, Wnt, transforming growth factor-β, Notch and vascular endothelial growth factor signaling pathways are expressed than in the adult liver transcriptome. In physiological contrast, only the BM-MSC and liver transcribed a statistically significant proportion of genes that are classified as part of the complement system (Supplementary Material, Fig. S3).

Quantitative transcript levels of individual components assigned to statistically over-represented signaling pathways are similar among the hNCC, hES3 and hES4 cells relative to the MSC, liver and lung (Fig. 7), as indicated by the cluster analysis (Fig. 6A). In fact, such pathways are functionally interconnected, as demonstrated by the expression of many intracellular components assigned to both integrin-related and IGF-related cascades (Fig. 8).

In summary, at multiple levels, the type and proportional representation of RNA transcripts in hNCC are most similar to the least committed cell types included in these analyses, the hES cell lines.

Neural crest-specific marker profile

To filter for those transcripts most specific to hNCC that might reside in postnatal tissues, we performed tissue preferential expression (TPE) analysis (39), using SAGE data from 14 other normal postnatal human cells or tissues, initially excluding hES cells. TPE is based both on the number of tissues in which a gene is present (range of expression) and its quantitative expression level in each. A score for each gene was plotted against the number of times that each gene was observed among the list of all tissues (Supplementary Material,

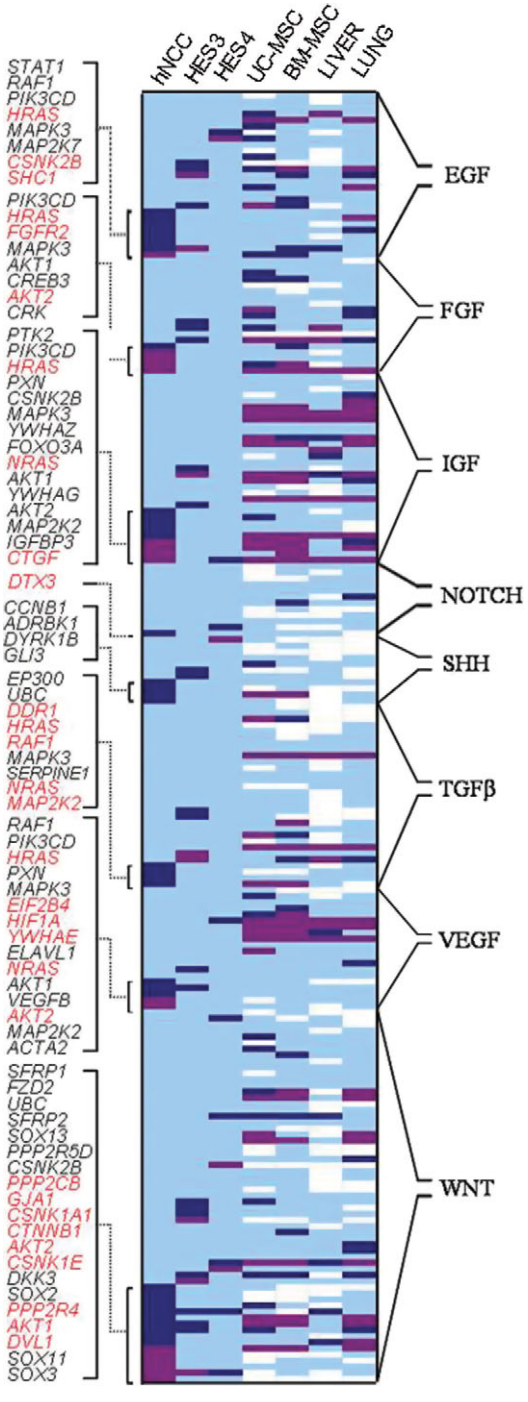

Figure 7. Growth factor pathways are similarly activated in hNCC and hES cells, relative to other cell types. Heat map representing expression levels in hNCC, hES3, hES4, UC-MSC, BM-MSC, adult liver and adult lung cells of lists of genes attributed by IP software to various signaling pathways. These include but are not exclusive to the epidermal growth factor (EGF), fibroblast growth factor (FGF), insulin-like growth factor (IGF), NOTCH, Sonic hedgehog (SHH), transforming growth factor (TGF)-beta, vascular endothelial growth factor (VEGF) and WNT pathways. Gene symbols for the more abundant hNCC tags are listed on the left, with high concomitant hES expression in red text. White, tags-per-million (tpm) value equal to 0; light blue, $1 \leq \text{tpm} \leq 49$; dark blue, $50 \leq \text{tpm} \leq 100$; magenta, tpm value >100.

Fig. S4). Functional annotation of the 1248 genes with TPE range = 15 correlated their ubiquitous expression with the finding that ~60% are structural ribosome constituents or proteins involved in primary cellular energy metabolism.

The hNCC most-specific gene list (range = 1) contained 119 transcripts (Supplementary Material, Table S8), many of which are involved in the regulation of transcription and DNA-histone packaging. Parallel analyses of hES3 or hES4, excluding the other hES cell line and hNCC, demonstrated that 48 of these 119 also had a TPE range = 1 in each. Thus, across the 16 distinct human cell and tissue types examined, 43 low-abundance transcripts were most specific to hNCC (Table 1) and 27 of these are also transcribed by C12–13 embryonic neuroepithelium (results not shown), in keeping with other expression data (Figs 3, 4, 6 and 9).

Of the 20 TPE = 1 genes that are currently annotated by Gene Ontology, half encode DNA-binding proteins such as DBX2 (40), ZNF157, HOXC5 and TIGD3. We confirmed hNCC expression of *HOXC5*, *C2ORF63 (FLJ31438)*, *ZNF417* and *AMIGO3* by RT–PCR in five distinct hNCC lines (Fig. 9A). Adult liver also transcribes *ZNF417* and *AMIGO3*, although transcript abundance was below the threshold of SAGE detection in a publicly accessible liver bank with 66861 short (10 bp) tags, accounting for the TPE result. *HOXC5* expression pattern in whole human embryo sections (Fig. 9B–G) correlates well with trunk-level neural tube expression (rhombomere 8 and caudal) noted in mouse embryos (41). Indeed, *HOXC5* is one of only three genes from this list whose orthologue is transcribed by mNCSC/NCP (22), although many of the other human transcripts are novel and do not yet have validated orthologues. *HOXC5* is expressed by hNCC entering the dorsal root, within the neural tube, in hypaxial muscle precursors (Fig. 9D) and in limb mesenchyme (Fig. 9B). Transcripts are more easily detected at C15 in the ventral motor horns, median ventricular zone and floorplate, dorsal root ganglia, as well as in the forming vertebral bodies, muscle precursors and stomach wall (Fig. 9F).

On the whole, these results show that hNCC can naturally be distinguished from hES cells by the expression of a highly specific and small subset of markers. However, like hES cells, hNCC keep many molecular characteristics of their *in vivo* embryonic phenotype, as well as the capacity for self-renewal *in vitro*.

DISCUSSION

The stem cell profile of hNCC

The shared transcriptional signature between hNCC and hES cells was startling. Trunk-level NCC in animals have been historically described as relatively oligopotent compared to the cephalic NCC. However, the extent and onset of trunk-level NCC lineage restriction remains controversial (42). *In vivo*, cartilage and bone, smooth muscle and adipocytes (among other derivatives) come either from the trunk mesoderm or the cephalic neural crest. Avian trunk-level NCC have the

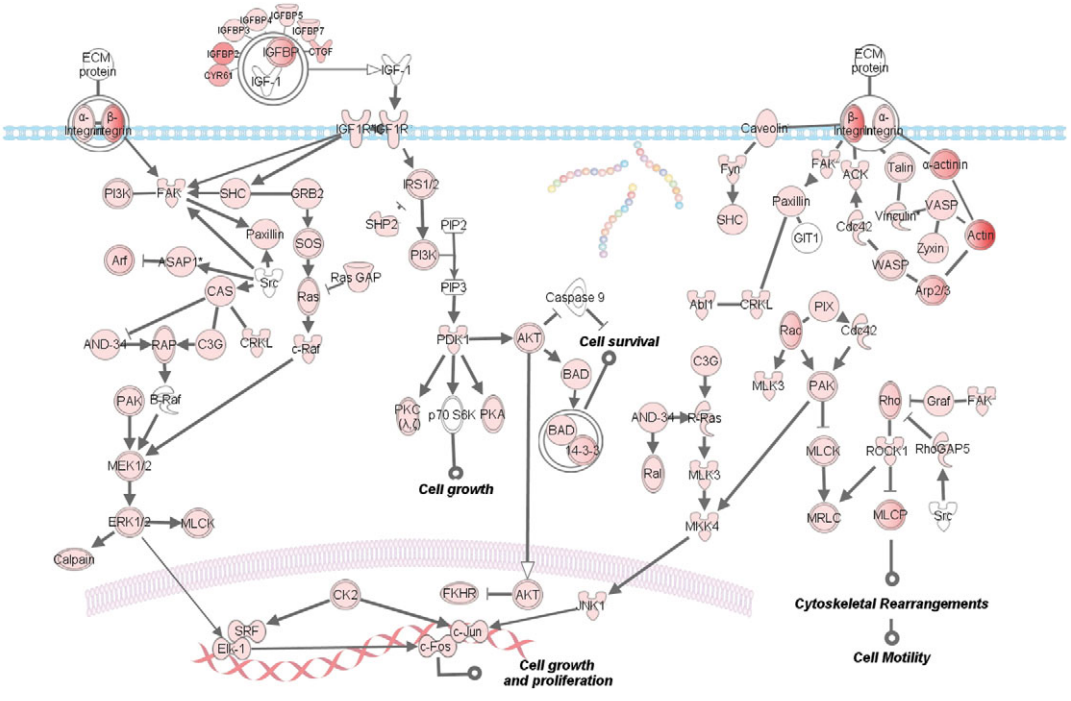

Figure 8. Schematic view of part of the IGF-1 and integrin signaling pathway genes transcribed by hNCC. Pink-tinted symbols are genes expressed by hNCC, and documented physical interactions exist for each edge depicted.

ability *in vitro* to not only give rise to cartilage and bone (3) and smooth muscle cells (43), but to adipocytes as well (44). Cartilage, bone, smooth muscle and adipocytes can of course also differentiate from BM-MSC, as many studies attest. Interestingly, Takashima *et al*. (13) have demonstrated that some murine BM-MSC in fact are seeded by neuroepithelial-derived NCC. We expected to find a closer relationship to BM-MSC or neural crest derivative transcriptomes than to hES lines.

Our data, therefore, have some provocative implications for the use of 'markers' in developmental biology. For example, like in hES lines, some early markers of definitive embryonic endoderm (*FOXA2*, *GPC1*, *TM4SF2* and *CXCR4*) are found in the hNCC bank, but not others (*SOX17* and *HHEX*). This hints that hNCC may have the potential to add new, non-physiological progeny to the already extensive list of *in vivo* derivatives; an overlapping contingent of these genes is expressed by the fully pluripotent hES3 line (31). However, under some initial conditions varying the physical matrix, seeding density and the proportions of exogenous growth factors (available upon request), we have not yet successfully oriented hNCC differentiation *in vitro*. A novel set of conditions that remains to be tested is a medium free specifically of blocks to β-catenin signaling and ERK activation, inhibitors of which are necessary and sufficient for mouse ES cell self-renewal (45).

The simultaneous transcription of *SOX2*, *NANOG* and *POU5F1* is not sufficient to confer ES identity, since we find all three in the hNCC lines. In addition, hNCC express *LIN28*, many members of the *KLF* gene family and *MYCN*, *MYCL1* and a number of MYC-binding proteins. Forced expression of various combinations of members of two of these gene families with *POU5F1* and *SOX2* have all led to pluripotent stem cell induction from human somatic cells (46,47).

Somatic cells from the pharyngula-stage embryo, such as these hNCC, may have a closer ground state to a pluripotent phenotype than do adult somatic cells, with their very low frequency of inducibility. With respect to 'pluripotency' markers, chicken embryonic neuroepithelium and neural folds express *Sox2* from stages preceding NCC formation (48). In contrast with our findings in human embryos, *in situ* hybridization does not demonstrate the presence of *Nanog* transcripts at neurula/pharyngula stages of mouse development (49). However, *Pou5f1* (*Oct3/4*), although best known for its expression by the murine germ-cell lineage and cell lines derived therefrom, is also expressed by both embryonic ectoderm and neuroepithelium (50), and it plays a permissive role in all germ layers in zebrafish (51). We interpret the similar molecular signature of hNCC and hES cells as an unprecedentedly detailed example of a homologous developmental process (52) underlying the state of multipotency. Further studies will be necessary to test the ability of hNCC *in vivo* to differentiate into functional cells representative of any or all tissue types with a teratoma assay; a better test would be their ability to contribute to chimeras, in particular to the germ line of such chimeras, before claims of pluripotency could be supported. However, for ethical and practical reasons, such experiments are impracticable by our laboratories.

Comparative embryology

In humans, as in other vertebrates, there is a lag in time between the peak migration windows for NCC from the cephalic versus

Table 1. Novel hNCC markers

Symbol	hNCC (tpm)	GO annotation Function	Process	Component	EST sources (NCBI) Embryo[a]	Foetus[b]	Adult[c]
AMIGO3	20	Protein binding	Cell adhesion	Membrane	0	1	7
C10ORF85	20	nd	nd	nd	0	3	1
DKFZP761N09121	20	Protein transporter activity	Intracellular vesicle-medicated transport	Golgi-associated vesicle	nd	nd	nd
DNAH1	20	Microtubule motor activity	Ciliary or flagellar motility	Axonemal dynein complex	2	7	84
FAM70B	20	nd	nd	Membrane	0	4	12
FLJ16139=DBX2	20	Transcription factor activity	Regulation of transcription, DNA-dependent		0	0	2
FLJ20345=MKS1	20	nd	nd	nd	5	19	52
FLJ31295=ZNF641	20	DNA binding, metal ion binding	Regulation of transcription, DNA-dependent	Intracellular	4	1	43
FLJ31438	40	nd	nd	nd	1	9	27
FLJ40126=C12orf40	20	nd	nd	nd	nd	nd	nd
GABRR3	20	GABA-A receptor activity	Synaptic transmission	Postsynaptic membrane	nd	nd	nd
HIST1H1D	20	DNA binding, protein binding	Nucleosome assembly, chromosome biogenesis	Nucleosome, chromosome	0	1	1
HIST1H2BE	20	nd	nd	nd	nd	nd	nd
HIST1H2BJ	20	DNA binding	Nucleosome assembly, chromosome organization	Nucleosome, chromosome	0	1	2
HIST1H3B	20	nd	nd	nd	nd	nd	nd
HIST1H4F	20	nd	nd	nd	nd	nd	nd
HMFN0672=C8orf80	40	nd	nd	nd	0	2	0
HOXC5	99	Transcription factor activity	Regulation of transcription from RNA pol II promoter	Nucleus	1	8	10
HOXD9	20	RNA pol ll transcription factor activity	Regulation of transcription, DNA-dependent	Nucleus	0	22	20
KCNG2	20	Potassium channel activity	Ion transport, potassium ion transport	Membrane	nd	nd	Nd
KIAA1822L	20	Catalytic activity	nd	nd	0	1	4
LOC255177	20	nd	nd	nd	0	0	0
LOC352909=C19orf51	20	nd	nd	nd	4	27	56
LOC400340	20	nd	nd	nd	nd	nd	nd
LOC401021	20	nd	nd	nd	nd	nd	nd
LOC401485	20	nd	nd	nd	nd	nd	nd
LOC440502	20	nd	nd	nd	nd	nd	nd
LOC440993	20	nd	nd	nd	nd	nd	nd
LOC441053	20	nd	nd	nd	0	23	177
LOC493860=CCDC73	20	nd	nd	nd	nd	nd	nd
MGC16372=C2orf39	20	nd	nd	nd	0	5	6
MGC48915=C1QTNF9	20	nd	Phosphate transport	Cytoplasm			
PP3856=NAPRT1	20	Nicotinate phosphoribosyltransferase	Pyridine nucleotide biosynthetic process	nd	1	9	64
PRDM12	59	DNA binding, metal ion binding	Regulation of transcription, DNA-dependent	Nucleus	0	2	0
PRH1	20	nd	nd	nd	1	4	353
RHEBL1	20	GTP binding	Small GTPase mediated signal transduction	Intracellular	0	0	6
SH3GLP2	20	nd	nd	nd	nd	nd	nd
STOX2	20	nd	nd	nd	nd	nd	nd
TIGD3	20	DNA binding	Regulation of transcription	Nucleus	0	5	4
UCN	20	Neuropeptide hormone activity	G-protein coupled receptor	Extracellular region	0	0	12
WNT7A	40	Receptor binding, signal transducer activity	Wnt receptor signaling pathway	Extracellular region	0	4	8
ZNF 157	20	DNA binding, metal ion binding	Regulation of transcription, DNA-dependent	Nucleus	0	0	0
ZNF417	40	DNA binding, metal ion binding	Regulation of transcription, DNA-dependent	Nucleus	1	4	13

Genes most specific to hNCC as shown by TPE analysis (range value = 1) and not expressed in hES cells.
[a]Count of human expressed sequence tags, out of 125 776 total from embryonic tissues.
[b]Count of human expressed sequence tags, out of 557 809 total from fetal tissues.
[c]Count of human expressed sequence tags, out of 1 899 694 total from postnatal tissues (UniGene's EST ProfileViewer).

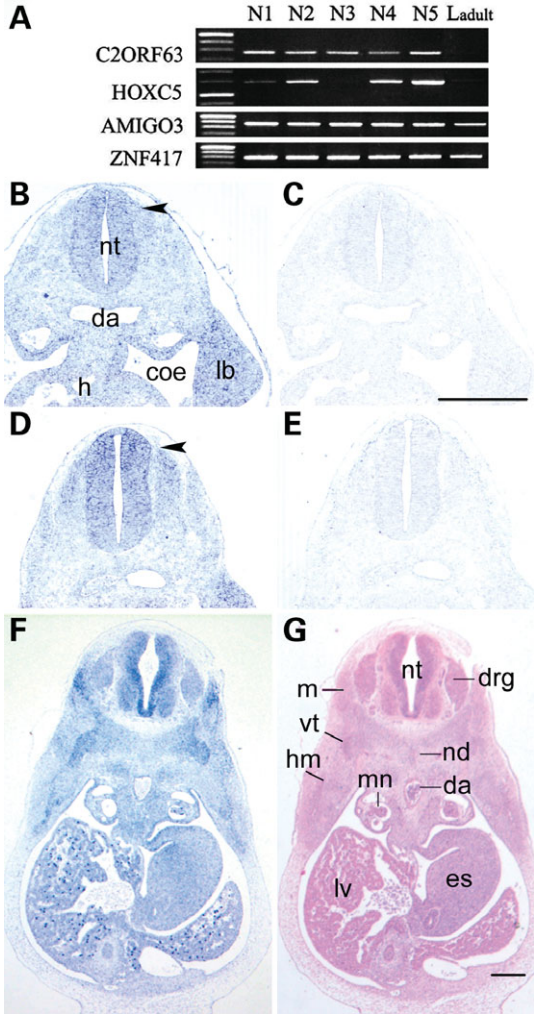

Figure 9. Prediction of mRNA subset specific to hNCC. (**A**) RT–PCR for four genes with TPE score=1 shows that it is possible to amplify *AMIGO3* and *ZNF417* from an adult human liver, although below the level of detection by a publicly available SAGE bank from another sample. *HOXC5* and *C2ORF63* were specific to most or all hNCC lines. (**B**) *In situ* hybridization to *HOXC5* demonstrates expression in the human embryo at Carnegie stage 13 (C13; cf. Fig. 4E) in most cells at the level of the limb bud, with higher levels in the neural tube, the neural crest (arrowhead), the limb mesenchyme and the rest of the somatopleural mesoderm. (**C** and **E**) Sense probe-hybridized adjacent transverse sections to previous frames demonstrate specificity of signal. (**D**) *HOXC5* in a different C13 embryo section taken at a more rostral transverse level than in (B); expression appears reduced in the floorplate relative to the rest of the neural tube. (**F**) At C15 (cf. Fig. 6E–G), *HOXC5* is expressed by the developing vertebra, muscle precursors, ventricular zone of the neural tube and motor horns; it is also transcribed by all other tissues in this section when compared with the sense-hybridized adjacent section (not shown). (**G**) Adjacent HE section. Thus, markers are only specific to a particular cell type within a given spatiotemporal context. Abbreviations: cœ, cœlom; da, dorsal aortae; h, heart; hm, hypaxial muscle precursors; lb, limb bud; lv, liver; m, muscle; nd, notochord; nt, neural tube; vt, vertebra.

the trunk levels. Species differences appear in the uncoupling of migration from that of neural tube closure. We observed that the majority of hNCC migrate after neural fold apposition along the body axis and continue to do so after fusion, similar to other amniotes and amphibia but in contrast to the fish (4). Recent anatomical observations imply that the first cranial hNCC begin migrating well before closure (53), with fusion catching up with hNCC delamination by the hindbrain. These two processes use common molecular mechanisms for the rearrangement of the actin cytoskeleton necessary for both convergent extension (54) and epithelio-mesenchymal transition (EMT)/migration (55), respectively.

Many gene transcripts found in avian or mouse migratory NCC are present in hNCC under their conditions of isolation. However, as many or more appeared to be unique to either mouse or to humans. A larger proportion still was expressed differentially to a statistically significant level. This is perhaps unsurprising considering the phenomenal plasticity of the neural crest lineage and its capacity to engender such diverse and species-specific structures as the turtle shell (56) and the deer antler (57). A greater total tag count in the hNCC bank may also have contributed to the detection of rarer transcripts relative to the mouse SAGE banks.

Among the human-specific expressed genes were a number that fell into characterized signaling pathways from other cellular systems, not necessarily suspected to have a role in the hNCC lineage. One phenotypic difference between human and animal NCC is that mouse cells are described to have a self-renewing potential limited to some weeks (22) while hNCC, cultivated without a feeder layer, can survive many months in our conditions without division arrest or large-scale differentiation. Animal NCC populations appear to be a heterogeneous collection of partially committed progenitors *in vitro*, be they of avian (16,58) or murine (15,59) origin. Henion and Weston (20) concluded that nearly half of the avian NCC from the trunk level was fate-restricted precursors generating a single cell type. Cluster analysis bears out our empirical observations by grouping the cultured hNCC transcriptome most closely with pluripotent embryonic stem cells, rather than with any committed neural crest progeny.

Novel hNCC markers

We have applied TPE to generate two detailed sub-profiles: that of highly specific hNCC markers and another of shared hNCC-hES molecules, which may be applied in the future to cell sorting. Currently, attempts to isolate NCC-like progenitors from human skin have relied on external phenotypic characteristics (10,60). Cell sorting has been used to enrich for rat NCC with tissue-restricted stem cell capacities; however, the marker employed was an elegant but artificial construct only applicable to genetically modified rodents (42). As identified, the rat cells had already become lineage-restricted in their peripheral niche.

Individual transcripts in the most-specific TPE list clearly do not signify NCC identity. The combination of all or a subset of them, and perhaps their quantitative levels, may however be sufficient to uniquely identify hNCC with a similar lack of fate commitment. One advantage of the SAGE approach is illustrated by the fact that 39% of unique tags in the hNCC library could not be reliably assigned to known mRNAs. Some of this ambiguity must be due to sequencing errors or polymorphisms (61), but a large subpopulation probably corresponds to novel transcripts expressed in hNCC. One indication is that 296 (1.3%) of these tags mapped readily to recently predicted natural antisense RNA (naRNA) sequences (62) (Supplementary

Material, Table S9). Others may correspond to primary microRNA transcripts. New membrane proteins will be particularly helpful for fluorescence-activated cell sorting experiments.

The results presented here will help in reliably isolating highly multipotent NCC from pre- or postnatal human tissues. These cells could be used for transplantation therapies in such diseases as peripheral demyelinating neuropathies or HSCR.

Candidate genes for human disease

As a point of comparison between early hNCC and their normal or pathological progeny, this hNCC signature can elucidate the molecular underpinnings of additional diseases. Many genes already known to cause human neurocristopathies are expressed in the hNCC SAGE library, including *PAX3* and *SOX10* (Waardenburg syndrome), *KIT* and *SNAI2* (piebaldism), *MSX1* (tooth agenesis/orofacial cleft), *CHD7* (CHARGE syndrome), and *ECE1*, *EDNRB*, *ITGB1*, *KIAA1279* and *ZEB2* (HSCR). Some neurocristopathies arise from the disruption of hNCC migration and/or the EMT from the neuroepithelium (14). Currently, a limited number of such EMT-associated disease genes include *EFNB1, FOXC1*, *LAMA5*, *SEMA3C* and *SNAI2* and are all known and expressed in this hNCC SAGE bank. The examination of functional modules such as EMT may be a fruitful strategy to identify new neurocristopathy genes. For example, the SNAG corepressor domain of SNAI2, found only in vertebrate homologues of Drosophila *snail*, has recently been implicated in *Xenopus* NCC specification within the neuroepithelium (63). Potential co-repressors *JUB*, *LIMD1*, *TRIP6, WTIP* and *ZYX* are all transcribed by hNCC, demonstrating the utility of such species- and tissue-specific expression banks for integrating dispersed functional data and generating new hypotheses for candidate genes to correlate with known disease loci. The use of our self-renewing hNCC lines for functional analyses of such candidate genes is a particularly appealing application of their potential in all senses of the word.

MATERIALS AND METHODS

Human embryos

Human embryos were collected from pregnancies legally terminated using the mefiprestone protocol, in concordance with French bioethics law 2004-800 and with the approval of the Necker hospital ethics committee.

hNCC isolation

Whole human embryos under 8 weeks' gestation were dissected away from their annexes in ice-cold phosphate-buffered saline (PBS). A piece of chorionic villus was treated immediately for 1 h in colchicine-DMEM before 10 min hypotonic shock in 0.075 M KCl and 1:3 acetic acid:ethanol fixation for three changes of 10 min each. Nuclei were spread, stained for G-bands and chromosomes counted according to standard karyotype procedure. Cephalic neural tube segments were isolated by cutting perpendicular to the long axis behind the optic vesicles and again at somite pair 5, and trunk-level neural tube segments from somite 5 to the last available somite pair. Other tissues were trimmed away and the segment placed in 6 mg/ml pancreatin (Sigma P3292) in PBS for 7 min. Pieces were transferred into clean PBS to tease away all tissues, including the tightly adherent notochord and ectoderm, from neural tubes. The enzymatic reaction was stopped and neural tubes maintained thereafter in complete hNCC medium.

Initial hNCC migration and replating on collagen I-coated plates (Becton Dickinson) used the following medium: Dulbecco's Modified Essential Medium/BGJb (Fitton-Jackson modified)/Ham's F12 [3:1:6 v/v, Sigma] supplemented with 12% complement-inactivated fetal calf serum (PAN Biotech); 1× penicillin/streptomycin (Invitrogen); 2 mM glutamine, 10 mM HEPES, 0.1 μg/ml hydrocortisone, 10 μg/ml transferrin, 0.4 ng/ml T3 (3,3,5-thio-iodo-thyronine), 10 pg/ml glucagon, 100 pg/ml epidermal growth factor, 1 ng/ml insulin and 200 pg/ml fibroblast growth factor 2 (Sigma). Tubes were transferred to the centre of 35 mm collagen-I dishes with 300 μl medium and placed in humidified CO_2 incubators at 37°C for 8 h to facilitate adhesion. One milliliter of medium was then added for a further 16 h, during which cells migrated onto the plastic. Tubes were removed with a fire-polished Pasteur pipette and primary cultures allowed to grow for another day before first replating. Passages were obtained by dissociation from plates using 3 min trypsin-EDTA treatment and complete hNCC medium to stop the reaction. No feeder layer cells or other adjuvants (e.g. leukemia inhibitory factor and embryo extract) were necessary for long-term maintenance for up to 9 months.

Karyotyping

Standard G-band karyotypes were performed on 10 metaphases from fresh trophoblast and thawed embryonic hNCC.

Immunocytochemistry

Cells were prepared for immunostaining by fixation in 4% paraformaldehyde for 20 min and subsequent permeabilization for 25 min with 0.1% Tween-20 in PBS. Primary antibodies: monoclonal anti-α-smooth muscle actin (Cy3-conjugated, Sigma C6198), monoclonal anti-neuron specific class III beta-tubulin (TuJ1) (R&D Systems), polyclonal anti-NCAM (Chemicon), polyclonal anti-tyrosine hydroxylase (Chemicon AB1542), polyclonal anti-calcitonin (Chemicon AB910), monoclonal anti-human glial filament acidic protein (GFAP) (Cymbus Biotechnology), polyclonal anti-SOX9 (Chemicon AB5535) and polyclonal anti-SOX2 (Abcam AB15830). Nuclei were counterstained (Hoechst).

Sage library construction and RT–PCR

Standard protocols (64,65) were used to construct a LongSAGE library from 5 μg total RNA (RNeasy Mini, Qiagen) prepared from trunk-level hNCC of a C13 embryo after 7 passages. In brief, *Nla*III and *MmeI* restriction enzymes (New England Biolabs) were used for tag generation. After 3 h self-ligation, 10′ at 65°C and 2′ on ice, purified concatemers were subsequently cloned into pZERO-1 (Invitrogen). Individual

clones were selected and sequenced. LongSAGE tags (17 bp) were extracted from raw sequence data using SAGE2000 version 4.5 Analysis Software. The LongSAGE tag database was linked to the SAGE Map database (NCBI) and the ACTG web tool (http://retina.med.harvard.edu/ACTG/) (66) for tag-to-gene mapping. The original SAGE data in this paper are available from the NCBI Gene Expression Omnibus (GEO) under accession number GSE8368.

Total RNA from five individual hNCC lines, C12 and C13 isolated neural tubes (four and two pooled, respectively) and four pooled C13 liver buds were prepared using the RNeasy Mini kit (Qiagen); total adult human liver RNA was obtained from Clontech. After reverse transcription (Applied Biosystems), PCR amplifications for gene validations were performed in a final volume of 25 μl using 50 ng of Nanodrop-quantified cDNA and 35 amplification cycles. For quantitative RT–PCR, LightCycler Fast Start DNA Master^{PLUS} SYBR green I (Roche) was used according to the manufacturer's protocol. *ACTB* was used to normalize data for the calculations of ΔCt. Primers are listed in Supplementary Material.

In situ hybridization

Intact, euploid human embryos were processed for *in situ* hybridization as described elsewhere (67). Primers used for DNA matrix synthesis by PCR are listed in Supplementary Material.

Statistical analyses

Hierarchical clustering was performed to create a multi-level binary cluster tree, linking tissue or cell types by similarity. In addition to the hNCC bank, data were obtained from 11 banks of the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/): GSM676 (brain white matter), GSM761 (cerebellum), GSM762 (lung), GSM708 (kidney), GSM785 (liver), GSM824 (muscle), GSM31931 (substantia nigra), GSM41359 (hES3), GSM41362 (hES4), GSM48250 (sciatic nerve), GSM48251 (Schwann cells *in vitro*); one from the Cancer Genome Anatomy Project http://cgap.nci.nih.gov/SAGE/SAGELibraryFinder:SAGE_Prostate_normal_B_2 (prostate); and for bone marrow and umbilical cord vein mesenchymal stem cells, data from the website http://bit.fmrp.usp.br/msc_tags/ (32) (current requests for raw data directly to authors). Counts from all SAGE libraries were normalized to a total of one million for input to Cluster 3.0 freeware (http://www.geo.vu.nl/~huik/cluster.htm) (68). Average distances were classed to link samples into clusters. In order to compare and validate the robustness of the results, Euclidean (squared), Pearson (uncentered) and Spearman rank numeric scales were all computed to calculate average distances. The dendogram was made with TreeView freeware (http://jtreeview.sourceforge.net/). Data were functionally annotated using Pathways Analysis 5.0 (Ingenuity Systems) and DAVID software (69).

The TPE analysis (39) was performed with the following additional samples (GEO or CGAP accession numbers): Bone marrow (GSM14784), placenta (GSM14749), muscle (GSM 824), liver (GSM785), stomach (SAGE_Stomach_normal_MD_13), kidney (GSM708), lung (GSM762), breast (SAGE_Breast_normal_myoepithelium_AP_myoepithelial1), prostate (SAGE_Prostate_normal_B_2), ovary (SAGE_Ovary_normal_CS_HOSE_4), endometrium (SAGE_Uterus_endometrium_normal_CS_DI1), brain white matter (GSM676), whole cerebellum (GSM761) and whole spinal cord (SAGE_Spinal_cord_normal_B_1).

For the detection of statistically differentially expressed genes, we used the IDEG6 web tool (http://telethon.bio.unipd.it/bioinfo/IDEG6_form/). The Fisher exact test (significance threshold 0.05) was used with a Bonferroni correction to account for multiple testing (29).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* Online.

FUNDING

This work was supported by the Avenir program of the Institut National de la Santé et de la Recherche Médicale, the Consortium National de Recherche en Génomique, the Association Française contre les Myopathies, and Agence National de la Recherche funding (MRAR2007 HIR-GENET). S.T., C.B., P.X. and M.S. are supported by grant NS039818 to M.S. from the US National Institutes of Health. Funding to pay the Open Access Charge was provided by the Fondation pour la Recherche Médicale.

ACKNOWLEDGEMENTS

We thank Dr. M. Téboul and the Service d'Orthogénie at the Hôpital Broussais, Paris, France, for their invaluable help. We are also grateful to Drs. J. Amiel, T. Attié-Bitach, O. Danos, E. Detrait, E. Dupin, E. Jabs and J.-P. Jais for their critical comments and to N. Brahimi, G. Goudefroye, G. Molina and C. Ozilou for their technical assistance.

Conflict of Interest statement. None declared.

REFERENCES

1. Abzhanov, A., Tzahor, E., Lassar, A.B. and Tabin, C.J. (2003) Dissimilar regulation of cell differentiation in mesencephalic (cranial) and sacral (trunk) neural crest cells in vitro. *Development*, **130**, 4567–4579.
2. Couly, G., Creuzet, S., Bennaceur, S., Vincent, C. and Le Douarin, N.M. (2002) Interactions between Hox-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebrate head. *Development*, **129**, 1061–1073.
3. McGonnell, I.M. and Graham, A. (2002) Trunk neural crest has skeletogenic potential. *Curr. Biol.*, **12**, 767–771.
4. Le Douarin, N. and Kalcheim, C. (1999) *The Neural Crest.* Cambridge University Press, Cambridge, UK.
5. Le Lievre, C.S. and Le Douarin, N.M. (1975) Mesenchymal derivatives of the neural crest: analysis of chimaeric quail and chick embryos. *J. Embryol. Exp. Morphol.*, **34**, 125–154.
6. Kruger, G.M., Mosher, J.T., Bixby, S., Joseph, N., Iwashita, T. and Morrison, S.J. (2002) Neural crest stem cells persist in the adult gut but undergo changes in self-renewal, neuronal subtype potential, and factor responsiveness. *Neuron*, **35**, 657–669.
7. Li, H.Y., Say, E.H. and Zhou, X.F. (2007) Isolation and characterization of neural crest progenitors from adult dorsal root Ganglia. *Stem Cells*, **25**, 2053–2065.

8. Fernandes, K.J., McKenzie, I.A., Mill, P., Smith, K.M., Akhavan, M., Barnabe-Heider, F., Biernaskie, J., Junek, A., Kobayashi, N.R., Toma, J.G. *et al.* (2004) A dermal niche for multipotent adult skin-derived precursor cells. *Nat. Cell Biol.*, **6**, 1082–1093.
9. Sieber-Blum, M., Grim, M., Hu, Y.F. and Szeder, V. (2004) Pluripotent neural crest stem cells in the adult hair follicle. *Dev. Dyn.*, **231**, 258–269.
10. Wong, C.E., Paratore, C., Dours-Zimmermann, M.T., Rochat, A., Pietri, T., Suter, U., Zimmermann, D.R., Dufour, S., Thiery, J.P., Meijer, D. *et al.* (2006) Neural crest-derived cells with stem cell features can be traced back to multiple lineages in the adult skin. *J. Cell Biol.*, **175**, 1005–1015.
11. Miura, M., Gronthos, S., Zhao, M., Lu, B., Fisher, L.W., Robey, P.G. and Shi, S. (2003) SHED: stem cells from human exfoliated deciduous teeth. *Proc. Natl. Acad. Sci. USA*, **100**, 5807–5812.
12. Motohashi, T., Aoki, H., Chiba, K., Yoshimura, N. and Kunisada, T. (2007) Multipotent cell fate of neural crest-like cells derived from embryonic stem cells. *Stem Cells*, **25**, 402–410.
13. Takashima, Y., Era, T., Nakao, K., Kondo, S., Kasuga, M., Smith, A.G. and Nishikawa, S. (2007) Neuroepithelial cells supply an initial transient wave of MSC differentiation. *Cell*, **129**, 1377–1388.
14. Etchevers, H.C., Amiel, J. and Lyonnet, S. (2006) Molecular bases of human neurocristopathies. *Adv. Exp. Med. Biol.*, **589**, 213–234.
15. Smith-Thomas, L.C. and Fawcett, J.W. (1989) Expression of Schwann cell markers by mammalian neural crest cells in vitro. *Development*, **105**, 251–262.
16. Baroffio, A., Dupin, E. and Le Douarin, N.M. (1988) Clone-forming ability and differentiation potential of migratory neural crest cells. *Proc. Natl. Acad. Sci. USA*, **85**, 5325–5329.
17. Spokony, R.F., Aoki, Y., Saint-Germain, N., Magner-Fink, E. and Saint-Jeannet, J.P. (2002) The transcription factor Sox9 is required for cranial neural crest development in Xenopus. *Development*, **129**, 421–432.
18. Uwanogho, D., Rex, M., Cartwright, E.J., Pearl, G., Healy, C., Scotting, P.J. and Sharpe, P.T. (1995) Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. *Mech. Dev.*, **49**, 23–36.
19. Baroffio, A. and Blot, M. (1992) Statistical evidence for a random commitment of pluripotent cephalic neural crest cells. *J. Cell Sci.*, **103**, 581–587.
20. Henion, P.D. and Weston, J.A. (1997) Timing and pattern of cell fate restrictions in the neural crest lineage. *Development*, **124**, 4351–4359.
21. Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W. and Velculescu, V.E. (2002) Using the transcriptome to annotate the genome. *Nat. Biotechnol.*, **20**, 508–512.
22. Hu, Y.F., Zhang, Z.J. and Sieber-Blum, M. (2006) An epidermal neural crest stem cell (EPI-NCSC) molecular signature. *Stem Cells*, **24**, 2692–2702.
23. Gammill, L.S. and Bronner-Fraser, M. (2002) Genomic analysis of neural crest induction. *Development*, **129**, 5731–5741.
24. Touraine, R.L., Attie-Bitach, T., Manceau, E., Korsch, E., Sarda, P., Pingault, V., Encha-Razavi, F., Pelet, A., Auge, J., Nivelon-Chevallier, A. *et al.* (2000) Neurological phenotype in Waardenburg syndrome type 4 correlates with novel SOX10 truncating mutations and expression in developing brain. *Am. J. Hum. Genet.*, **66**, 1496–1503.
25. Espinosa-Parrilla, Y., Amiel, J., Auge, J., Encha-Razavi, F., Munnich, A., Lyonnet, S., Vekemans, M. and Attie-Bitach, T. (2002) Expression of the SMADIP1 gene during early human development. *Mech. Dev.*, **114**, 187–191.
26. Sanlaville, D., Etchevers, H.C., Gonzales, M., Martinovic, J., Clement-Ziza, M., Delezoide, A.L., Aubry, M.C., Pelet, A., Chemouny, S., Cruaud, C. *et al.* (2006) Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human development. *J. Med. Genet.*, **43**, 211–217.
27. Pinson, L., Auge, J., Audollent, S., Mattei, G., Etchevers, H., Gigarel, N., Razavi, F., Lacombe, D., Odent, S., Le Merrer, M. *et al.* (2004) Embryonic expression of the human MID1 gene and its mutations in Opitz syndrome. *J. Med. Genet.*, **41**, 381–386.
28. Martinovic-Bouriel, J., Bernabe-Dupont, C., Golzio, C., Grattagliano-Bessieres, B., Malan, V., Bonniere, M., Esculpavit, C., Fallet-Bianco, C., Mirlesse, V., Le Bidois, J. *et al.* (2007) Matthew-Wood syndrome: report of two new cases supporting autosomal recessive inheritance and exclusion of FGF10 and FGFR2. *Am. J. Med. Genet. A*, **143**, 219–228.
29. Romualdi, C., Bortoluzzi, S., D'Alessi, F. and Danieli, G.A. (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol. Genomics*, **12**, 159–162.
30. de Jonge, R.R., Vreijling, J.P., Meintjes, A., Kwa, M.S., van Kampen, A.H., van Schaik, I.N. and Baas, F. (2003) Transcriptional profile of the human peripheral nervous system by serial analysis of gene expression. *Genomics*, **82**, 97–108.
31. Richards, M., Tan, S.P., Tan, J.H., Chan, W.K. and Bongso, A. (2004) The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells*, **22**, 51–64.
32. Panepucci, R.A., Siufi, J.L., Silva, W.A., Jr, Proto-Siquiera, R., Neder, L., Orellana, M., Rocha, V., Covas, D.T. and Zago, M.A. (2004) Comparison of gene expression of umbilical cord vein and bone marrow-derived mesenchymal stem cells. *Stem Cells*, **22**, 1263–1278.
33. Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. *et al.* (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.*, **38**, 431–440.
34. Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A. *et al.* (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.*, **9**, 625–635.
35. Zaehres, H., Lensch, M.W., Daheron, L., Stewart, S.A., Itskovitz-Eldor, J. and Daley, G.Q. (2005) High-efficiency RNA interference in human embryonic stem cells. *Stem Cells*, **23**, 299–305.
36. Walker, E., Ohishi, M., Davey, R.E., Zhang, W., Cassar, P.A., Tanaka, T.S., Der, S.D., Morris, Q., Hughes, T.R., Zandstra, P.W. and Stanford, W.L. (2007) Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. *Cell Stem Cell*, **1**, 71–86.
37. Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
38. Skottman, H., Mikkola, M., Lundin, K., Olsson, C., Stromberg, A.M., Tuuri, T., Otonkoski, T., Hovatta, O. and Lahesmaa, R. (2005) Gene expression signatures of seven individual human embryonic stem cell lines. *Stem Cells*, **23**, 1343–1356.
39. Moreno, J.C., Pauws, E., van Kampen, A.H., Jedlickova, M., de Vijlder, J.J. and Ris-Stalpers, C. (2001) Cloning of tissue-specific genes using serial analysis of gene expression and a novel computational substraction approach. *Genomics*, **75**, 70–76.
40. Shoji, H., Ito, T., Wakamatsu, Y., Hayasaka, N., Ohsaki, K., Oyanagi, M., Kominami, R., Kondoh, H. and Takahashi, N. (1996) Regionalized expression of the Dbx family homeobox genes in the embryonic CNS of the mouse. *Mech. Dev.*, **56**, 25–39.
41. Geada, A.M., Gaunt, S.J., Azzawi, M., Shimeld, S.M., Pearce, J. and Sharpe, P.T. (1992) Sequence and embryonic expression of the murine *Hox-3.5* gene. *Development*, **116**, 497–506.
42. Mosher, J.T., Yeager, K.J., Kruger, G.M., Joseph, N.M., Hutchin, M.E., Dlugosz, A.A. and Morrison, S.J. (2007) Intrinsic differences among spatially distinct neural crest stem cells in terms of migratory properties, fate determination, and ability to colonize the enteric nervous system. *Dev. Biol.*, **303**, 1–15.
43. Real, C., Glavieux-Pardanaud, C., Vaigot, P., Le Douarin, N. and Dupin, E. (2005) The instability of the neural crest phenotypes: Schwann cells can differentiate into myofibroblasts. *Int. J. Dev. Biol.*, **49**, 151–159.
44. Billon, N., Iannarelli, P., Monteiro, M.C., Glavieux-Pardanaud, C., Richardson, W.D., Kessaris, N., Dani, C. and Dupin, E. (2007) The generation of adipocytes by the neural crest. *Development*, **134**, 2283–2292.
45. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.
46. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861–872.
47. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R. *et al.* (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–1920.
48. Rex, M., Orme, A., Uwanogho, D., Tointon, K., Wigmore, P.M., Sharpe, P.T. and Scotting, P.J. (1997) Dynamic expression of chicken *Sox2* and

Sox3 genes in ectoderm induced to form neural tissue. *Dev. Dyn.*, **209**, 323–332.
49. Hart, A.H., Hartley, L., Ibrahim, M. and Robb, L. (2004) Identification, cloning and expression analysis of the pluripotency promoting *Nanog* genes in mouse and human. *Dev. Dyn.*, **230**, 187–198.
50. Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. and Gruss, P. (1990) Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.*, **9**, 2185–2195.
51. Reim, G., Mizoguchi, T., Stainier, D.Y., Kikuchi, Y. and Brand, M. (2004) The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. *Dev.Cell*, **6**, 91–101.
52. Gilbert, S.F. and Bolker, J.A. (2001) Homologies of process and modular elements of embryonic construction. *J. Exp. Zool.*, **291**, 1–12.
53. O'Rahilly, R. and Muller, F. (2007) The development of the neural crest in the human. *J. Anat.*, **211**, 335–351.
54. Copp, A.J., Greene, N.D. and Murdoch, J.N. (2003) The genetic basis of mammalian neurulation. *Nat. Rev. Genet.*, **4**, 784–793.
55. De Calisto, J., Araya, C., Marchant, L., Riaz, C.F. and Mayor, R. (2005) Essential role of non-canonical Wnt signalling in neural crest migration. *Development*, **132**, 2587–2597.
56. Cebra-Thomas, J.A., Betters, E., Yin, M., Plafkin, C., McDow, K. and Gilbert, S.F. (2007) Evidence that a late-emerging population of trunk neural crest cells forms the plastron bones in the turtle Trachemys scripta. *Evol. Dev.*, **9**, 267–277.
57. Allen, S.P., Maden, M. and Price, J.S. (2002) A role for retinoic acid in regulating the regeneration of deer antlers. *Dev. Biol.*, **251**, 409–423.
58. Sieber-Blum, M. and Cohen, A.M. (1980) Clonal analysis of quail neural crest cells: they are pluripotent and differentiate in vitro in the absence of noncrest cells. *Dev. Biol.*, **80**, 96–106.
59. Stainier, D.Y., Bilder, D.H. and Gilbert, W. (1991) The B30 ganglioside is a cell surface marker for neural crest-derived neurons in the developing mouse. *Dev. Biol.*, **144**, 177–188.
60. Fernandes, K.J., Kobayashi, N.R., Gallagher, C.J., Barnabe-Heider, F., Aumont, A., Kaplan, D.R. and Miller, F.D. (2006) Analysis of the neurogenic potential of multipotent skin-derived precursors. *Exp. Neurol.*, **201**, 32–48.
61. Keime, C., Semon, M., Mouchiroud, D., Duret, L. and Gandrillon, O. (2007) Unexpected observations after mapping LongSAGE tags to the human genome. *BMC Bioinformatics*, **8**, 154.
62. Ge, X., Wu, Q., Jung, Y.C., Chen, J. and Wang, S.M. (2006) A large quantity of novel human antisense transcripts detected by LongSAGE. *Bioinformatics*, **22**, 2475–2479.
63. Langer, E.M., Feng, Y., Zhaoyuan, H., Rauscher, F.J., III, Kroll, K.L. and Longmore, G.D. (2008) Ajuba LIM proteins are snail/slug corepressors required for neural crest development in xenopus. *Dev. Cell*, **14**, 424–436.
64. Basrai, M.A. and Hieter, P. (2002) Transcriptome analysis of Saccharomyces cerevisiae using serial analysis of gene expression. *Methods Enzymol.*, **350**, 414–444.
65. Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.
66. Galante, P.A., Trimarchi, J., Cepko, C.L., de Souza, S.J., Ohno-Machado, L. and Kuo, W.P. (2007) Automatic correspondence of tags and genes (ACTG): a tool for the analysis of SAGE, MPSS and SBS data. *Bioinformatics*, **23**, 903–905.
67. Delous, M., Baala, L., Salomon, R., Laclef, C., Vierkotten, J., Tory, K., Golzio, C., Lacoste, T., Besse, L., Ozilou, C. *et al.* (2007) The ciliary gene RPGRIP1L is mutated in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome. *Nat. Genet.*, **39**, 875–881.
68. de Hoon, M.J., Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics.*, **20**, 1453–1454.
69. Dennis, G., Jr, Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.*, **4**, 3.