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

Sophie Thomas¹, Marie Thomas¹, Patrick Wincker², Candice Babarit¹, Puting Xu³, Marcy C. Speer³, Arnold Munnich¹,4, Stanislas Lyonnet¹,4, Michel Vekemans¹,4 and Heather C. Etchevers¹,5,*

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

†Died August 4, 2007.
/C3
To whom correspondence should be addressed. Tel: +33 562744587; Fax: +33 562744558; Email: heather.etchevers@inserm.fr

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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.

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).
A 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 substantially 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 passed 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, MSXI, SNAI2, SOX9, SOX10 and TWIST, as well as signaling molecules or membrane-bound receptors such as RET, GJA1, EDN1, 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 MSXI 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-
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 MIDI (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...
scripts identified in an early survey of avian NCC cDNAs (23), 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, HESLS, 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

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),
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...
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,
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
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>
<thead>
<tr>
<th>Symbol</th>
<th>GO annotation</th>
<th>Process</th>
<th>Component</th>
<th>EST sources (NCBI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMIGO3</td>
<td>Protein binding</td>
<td>Cell adhesion</td>
<td>Membrane</td>
<td>Embryo: 0, Foetus: 3, Adult: 7</td>
</tr>
<tr>
<td>C100RF85</td>
<td>Protein transporter activity</td>
<td>Intracellular vesicle-mediated transport</td>
<td>Golgi-associated vesicle</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>DKKFZP761N09121</td>
<td></td>
<td>Ciliary or flagellar motility</td>
<td>Axonemal-associated complex</td>
<td>2, 7, 84</td>
</tr>
<tr>
<td>DNAH1</td>
<td>Microtubule motor activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAM70B</td>
<td>Transcription factor activity</td>
<td>nd</td>
<td>nd</td>
<td>0, 4, 12</td>
</tr>
<tr>
<td>FLJ16139=DBX2</td>
<td></td>
<td>Regulation of transcription, DNA-dependent</td>
<td>Membrane</td>
<td>0, 0, 2</td>
</tr>
<tr>
<td>FLJ20345=MKS1</td>
<td>DNA binding, metal ion binding</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>nd</td>
<td>5, 19, 52</td>
</tr>
<tr>
<td>FLJ31295=ZNFG41</td>
<td></td>
<td></td>
<td>Intracellular</td>
<td>4, 1, 43</td>
</tr>
<tr>
<td>FLJ31438</td>
<td>GABA-A receptor activity</td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>FLJ40126=C12orf40</td>
<td></td>
<td>Synaptic transmission</td>
<td>Postsynaptic membrane</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>GABRR3</td>
<td>DNA binding, protein binding</td>
<td>Nucleosome assembly, chromosome biogenesis</td>
<td>Nucleosome, chromosome</td>
<td>0, 1, 1</td>
</tr>
<tr>
<td>HIST1H1D</td>
<td>DNA binding</td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>HIST1H2BE</td>
<td>DNA binding</td>
<td>Nucleosome assembly, chromosome organization</td>
<td>Nucleosome, chromosome</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>HIST1H2BJ</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>0, 0, 0</td>
</tr>
<tr>
<td>HIST1H3B</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>4, 27, 56</td>
</tr>
<tr>
<td>HIST1H4F</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>HMNF0672=C8orf80</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>HOXC5</td>
<td>Transcription factor activity</td>
<td>Regulation of transcription from RNA pol II promoter</td>
<td>Nucleus</td>
<td>1, 8, 10</td>
</tr>
<tr>
<td>HOXD9</td>
<td>RNA pol II transcription factor activity</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>Nucleus</td>
<td>0, 22, 20</td>
</tr>
<tr>
<td>KCNG2</td>
<td>Potassium channel activity</td>
<td>Ion transport, potassium ion transport</td>
<td>nd</td>
<td>nd, nd, 4</td>
</tr>
<tr>
<td>KIAA1822L</td>
<td>Catalytic activity</td>
<td>nd</td>
<td>nd</td>
<td>0, 1, 4</td>
</tr>
<tr>
<td>LOC255177</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>0, 0, 0</td>
</tr>
<tr>
<td>LOC352909=C19orf51</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>4, 27, 56</td>
</tr>
<tr>
<td>LOC400340</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>LOC401021</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>LOC401485</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>LOC440993</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>LOC441053</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>0, 23, 177</td>
</tr>
<tr>
<td>LOC493860=CCDC73</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>0, 5, 6</td>
</tr>
<tr>
<td>MGC16372=C2orf39</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>0, 5, 6</td>
</tr>
<tr>
<td>MGC48915=C1QTNF9</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>0, 5, 6</td>
</tr>
<tr>
<td>PPI358=NAPRT1</td>
<td></td>
<td>Nicotinate</td>
<td>nd</td>
<td>1, 9, 64</td>
</tr>
<tr>
<td>PRDM12</td>
<td></td>
<td>Nicotinate</td>
<td>nd</td>
<td>0, 2, 0</td>
</tr>
<tr>
<td>PRH1</td>
<td>GTP binding</td>
<td>nd</td>
<td>nd</td>
<td>1, 4, 353</td>
</tr>
<tr>
<td>RHEBL1</td>
<td></td>
<td>Small GTPase mediated signal transduction</td>
<td>Intracellular</td>
<td>0, 0, 6</td>
</tr>
<tr>
<td>SH3GLP2</td>
<td></td>
<td>nd</td>
<td>nd</td>
<td>nd, nd, nd</td>
</tr>
<tr>
<td>STOX2</td>
<td>DNA binding</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>Nucleus</td>
<td>0, 5, 4</td>
</tr>
<tr>
<td>TIGD3</td>
<td>Neuropeptide hormone activity</td>
<td>G-protein coupled receptor</td>
<td>Extracellular region</td>
<td>0, 0, 12</td>
</tr>
<tr>
<td>UCN</td>
<td>Receptor binding, signal transducer activity</td>
<td>Wnt receptor signaling pathway</td>
<td>Extracellular region</td>
<td>0, 4, 8</td>
</tr>
<tr>
<td>WNT7A</td>
<td>DNA binding, metal ion binding</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>Nucleus</td>
<td>0, 0, 0</td>
</tr>
<tr>
<td>ZNF 157</td>
<td>DNA binding, metal ion binding</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>Nucleus</td>
<td>1, 4, 13</td>
</tr>
<tr>
<td>ZNF417</td>
<td>DNA binding, metal ion binding</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>Nucleus</td>
<td>1, 4, 13</td>
</tr>
</tbody>
</table>

Genes most specific to hNCC as shown by TPE analysis (range value = 1) and not expressed in hES cells.

*Count of human expressed sequence tags, out of 125,776 total from embryonic tissues.

Count of human expressed sequence tags, out of 557,809 total from fetal tissues.

*Count of human expressed sequence tags, out of 1,899,694 total from postnatal tissues (UniGene’s EST ProfileViewer).
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), MSXI (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 neurocristopathies 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, WTIIP 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 mifepristone 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 x 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 CO2 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-a-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, NlaIII and Mmel 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), GSM768 (kidney), GSM824 (muscle), GSM31931 (substantia nigra), GSM41359 (heES3), GSM41362 (heES4), GSM48250 (somatic 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.fm.mbrp.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/~huijk/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://treetview.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_myoepeithelium_AP_myoeptiheli1), prostate (SAGE_Prostate_normal_B_2), ovary (SAGE_Ovary_normal_CS_HOSE_4), endometrium (SAGE_Uterus_endometrium_normal_CS_D11), 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/ibioinfo/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**


